

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:58:49 ; Search time 52 Seconds
(without alignments)

532.844 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGEELFTAVVPIVLELDG.....VLEFVTAAGTTHGMDELK 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 11641973 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCR_NEM_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/CTUS_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1270	100.0	238	9	US-09-575-847-2
2	1270	100.0	238	9	US-10-071-976-2
3	1246	98.1	238	9	US-10-057-505-2
4	1246	98.1	238	10	US-09-884-681-2
5	1246	98.1	238	12	US-10-024-686-2
6	1242	97.8	238	9	US-09-866-538-2
7	1242	97.8	238	9	US-09-900-345A-125
8	1242	97.8	238	9	US-10-121-258-10
9	1242	97.8	238	9	US-10-121-258-10
10	1242	97.8	238	10	US-09-920-922-4
11	1242	97.8	238	10	US-09-852-000-1
12	1239	97.6	243	9	US-09-900-345A-60
13	1239	97.6	243	9	US-09-900-345A-62
14	1239	97.6	243	9	US-09-900-345A-64
15	1239	97.6	243	9	US-09-900-345A-66
16	1239	97.6	243	9	US-09-900-345A-68
17	1239	97.6	243	9	US-09-900-345A-70
18	1239	97.6	1070	9	US-10-001-486B-2
19	1239	97.6	1099	9	US-10-259-864-4

20	1239	97.6	1147	9	US-10-259-864-1	Sequence 1, Appli
21	1239	97.6	1452	12	US-10-050-673-2	Sequence 2, Appli
22	1238	97.5	243	9	US-09-900-345A-54	Sequence 54, Appli
23	1238	97.5	243	9	US-09-900-345A-56	Sequence 56, Appli
24	1238	97.5	243	9	US-09-900-345A-58	Sequence 58, Appli
25	1238	97.5	243	9	US-09-900-345A-114	Sequence 114, App
26	1238	97.5	243	9	US-09-900-345A-116	Sequence 116, App
27	1238	97.5	243	9	US-09-900-345A-118	Sequence 118, App
28	1238	97.5	243	9	US-09-900-345A-120	Sequence 120, App
29	1237	97.4	243	9	US-09-900-345A-2	Sequence 2, Appli
30	1237	97.4	243	9	US-09-900-345A-4	Sequence 4, Appli
31	1237	97.4	243	9	US-09-900-345A-6	Sequence 6, Appli
32	1237	97.4	243	9	US-09-900-345A-8	Sequence 8, Appli
33	1237	97.4	243	9	US-09-900-345A-10	Sequence 10, Appli
34	1237	97.4	243	9	US-09-900-345A-12	Sequence 12, Appli
35	1237	97.4	243	9	US-09-900-345A-14	Sequence 14, Appli
36	1237	97.4	243	9	US-09-900-345A-16	Sequence 16, Appli
37	1237	97.4	243	9	US-09-900-345A-18	Sequence 18, Appli
38	1237	97.4	243	9	US-09-900-345A-20	Sequence 20, Appli
39	1237	97.4	243	9	US-09-900-345A-22	Sequence 22, Appli
40	1237	97.4	243	9	US-09-900-345A-24	Sequence 24, Appli
41	1237	97.4	243	9	US-09-900-345A-26	Sequence 26, Appli
42	1237	97.4	243	9	US-09-900-345A-28	Sequence 28, Appli
43	1237	97.4	243	9	US-09-900-345A-30	Sequence 30, Appli
44	1237	97.4	243	9	US-09-900-345A-32	Sequence 32, Appli
45	1237	97.4	243	9	US-09-900-345A-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1
US-09-575-847-2
; Sequence 2, Application US/09575847
; Publication No. US20030013149A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: WACHTER, Rebekka
; APPLICANT: REMINGTON, James
; TITLE OF INVENTION: LONG WAVELENGTH ENGINEERED FLUORESCENT PROTEINS
; FILE REFERENCE: RESEN1250-5
; CURRENT APPLICATION NUMBER: US/09/575, 847
; CURRENT FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 08/974,737
; PRIOR FILING DATE: 1997-11-19
; PRIOR APPLICATION NUMBER: US 08/911,825
; PRIOR FILING DATE: 1997-08-15
; PRIOR APPLICATION NUMBER: US 08/706,408
; PRIOR FILING DATE: 1996-08-30
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea
US-09-575-847-2

Query Match 100.0%; Score 1270; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.3e-113;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MSKGEELTAVVPIVLELDGNGKHFVSFGEGDVTYKLTLEFCTTGKLPVWPPTL 60
1 MSKGEELTAVVPIVLELDGNGKHFVSFGEGDVTYKLTLEFCTTGKLPVWPPTL 60
61 VTTFSGYQVCSRYPDHMKRHFDFKSAPEGVQRTLFFKDDGNYKTRAVEKSGDTLV 120
61 VTTFSGYQVCSRYPDHMKRHFDFKSAPEGVQRTLFFKDDGNYKTRAVEKSGDTLV 120
121 NRIELKGDIFKEDNGIILHKLLEYNNYSNIVYIMADKONGIKVNRKIRHNIEDSGVOLAD 180
121 NRIELKGDIFKEDNGIILHKLLEYNNYSNIVYIMADKONGIKVNRKIRHNIEDSGVOLAD 180
121 NRIELKGDIFKEDNGIILHKLLEYNNYSNIVYIMADKONGIKVNRKIRHNIEDSGVOLAD 180

QY 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 2

US-10-071-976-2
; Sequence 2, Application US/10071976
; Publication No. US20030036178A1
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; PROTEINS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/071,976
; FILING DATE: 05-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/465,142
; FILING DATE: 1999-DEC-16
; APPLICATION NUMBER: 08/974,737
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-071-976-2

Query Match 100.0%; Score 1270; DB 9; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.3e-113;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLDGVDNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVLDGVDNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
QY 61 VTTFSTGVQCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTTFSTGVQCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRTELKIDPKEDGNIIGHKLEYNYNNSHNYIMADKONGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDPKEDGNIIGHKLEYNYNNSHNYIMADKONGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 3

US-10-057-505-2
; Sequence 2, Application US/10057505
; Patent No. US20020164674A1
; GENERAL INFORMATION:
; APPLICANT: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA
; APPLICANT: AURORA BIOSCIENCES CORPORATION
; APPLICANT: TSJEN, Roger
; APPLICANT: HEIM, Roger
; APPLICANT: CUBITT, Andrew
; TITLE OF INVENTION: TANDEM FLUORESCENT PROTEIN CONSTRUCTS
; FILE REFERENCE: REGEN1260-3
; CURRENT APPLICATION NUMBER: US/10/057,505
; PRIOR FILING DATE: 2002-01-25
; PRIOR APPLICATION NUMBER: US 08/792,553
; PRIOR FILING DATE: 1997-01-31
; PRIOR APPLICATION NUMBER: US 09/396,003
; PRIOR FILING DATE: 1999-09-13
; NUMBER OF SEQ ID NOS: 31
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-057-505-2

Query Match 98.1%; Score 1246; DB 9; Length 238;
Best Local Similarity 97.9%; Pred. No. 2.5e-111;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLDGVDNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVLDGVDNGHKFSVSGEGDVTYGLKLFICTTGKLPVMPPTL 60
QY 61 VTTFSTGVQCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVKEGDTLV 120
Db 61 VTTFSTGVQCFSRPDMKRRHDFPKSAMPEGYVOORTIFFKDDGNYKTRAEVKEGDTLV 120
QY 121 NRTELKIDPKEDGNIIGHKLEYNYNNSHNYIMADKONGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRTELKIDPKEDGNIIGHKLEYNYNNSHNYIMADKONGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238
Db 181 YGQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGHMDELYK 238

RESULT 4

US-09-884-681-2
; Sequence 2, Application US/09884681
; Patent No. US20020061546A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.

; TITLE OF INVENTION: Assays for Protein Kinases Using
; Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/884,681
; FILING DATE: 19-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/679,865
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Storella, John S.
REGISTRATION NUMBER: 32,944
REFERENCE/DOCKET NUMBER: 023072-069000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-884-681-2

Query Match 98.1%; Score 1246; DB 10; Length 238;
Best Local Similarity 97.9%; Pred. No. 2.5e-111;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIVLELDGVNGHKFSVSGEGEDVYTGKLTLCFTTGTGKLPVWPPTL 60
DB 1 MSKGEELTGVVPIVLELDGVNGHKFSVSGEGEDATYGLTLKFTCTTGKLPVWPPTL 60
QY 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOORTIFFKDDGNYKTRAEEKFEGDLY 120
DB 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEEKFEGDLY 120
QY 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLIDGPVLLPDNHYLSTOSALS KDPNEKRDHVLLEFVTAAGITTHGMDELK 238
DB 181 YVQONTPLIDGPVLLPDNHYLSTOSALS KDPNEKRDHVLLEFVTAAGITTHGMDELK 238

RESULT 5

US-10-024-686-2
Sequence 2, Application US/10024686
Patent No. US20020123113A1
GENERAL INFORMATION:
APPLICANT: Tsien, Roger Y.
TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/024,686
FILING DATE: 17-Dec-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/057,995
FILING DATE: <Unknown>
APPLICATION NUMBER: 08/727,452
FILING DATE: 10-OCT-1996
APPLICATION NUMBER: US95/14692
FILING DATE: 13-NOV-1995
APPLICATION NUMBER: 08/337,915
FILING DATE: 10-NOV-1994
ATTORNEY/AGENT INFORMATION:

NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07257/032002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 238 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-024-686-2

Query Match 98.1%; Score 1246; DB 12; Length 238;
Best Local Similarity 97.9%; Pred. No. 2.5e-111;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIVLELDGVNGHKFSVSGEGEDVYTGKLTLCFTTGTGKLPVWPPTL 60
DB 1 MSKGEELTGVVPIVLELDGVNGHKFSVSGEGEDATYGLTLKFTCTTGKLPVWPPTL 60
QY 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOORTIFFKDDGNYKTRAEEKFEGDLY 120
DB 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEEKFEGDLY 120
QY 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPLIDGPVLLPDNHYLSTOSALS KDPNEKRDHVLLEFVTAAGITTHGMDELK 238
DB 181 YVQONTPLIDGPVLLPDNHYLSTOSALS KDPNEKRDHVLLEFVTAAGITTHGMDELK 238

RESULT 6

US-09-866-538-2
Sequence 2, Application US/09866538
Publication No. US20030032088A1
GENERAL INFORMATION:
APPLICANT: REGENTS OF THE UNIVERSITY OF CALIFORNIA
APPLICANT: TSJEN, ROGER
APPLICANT: CAMPBELL, ROBERT
TITLE OF INVENTION: NON-OLIGOMERIZING FLUORESCENT PROTEINS
FILE REFERENCE: REGEN1530-2
CURRENT APPLICATION NUMBER: US/09/866,538
CURRENT FILING DATE: 2001-05-24
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
US-09-866-538-2

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIVLELDGVNGHKFSVSGEGEDVYTGKLTLCFTTGTGKLPVWPPTL 60
DB 1 MSKGEELTGVVPIVLELDGVNGHKFSVSGEGEDATYGLTLKFTCTTGKLPVWPPTL 60
QY 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOORTIFFKDDGNYKTRAEEKFEGDLY 120
DB 61 VTFESYGYOCFSRYPDHMKRHDFFKSAMPEGYVOERTIFFKDDGNYKTRAEEKFEGDLY 120
QY 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKIDFKEDGNILGHKLEYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 7

US-09-900-345A-125
; Sequence 125, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-50S
; CURRENT APPLICATION NUMBER: US/09/900.345A
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU P8078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 125
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: GFP humanized
US-09-900-345A-125

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-11;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
QY 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
Db 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 8

US-10-121-258-10
; Sequence 10, Application US/10121258
; Publication No. US20030059835A1
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger
; APPLICANT: Campbell, Robert
; TITLE OF INVENTION: MONOMERIC AND DIMERIC FLUORESCENT
; FILE REFERENCE: UC083.1CP2CPI
; CURRENT APPLICATION NUMBER: US/10/121.258
; PRIOR FILING DATE: 2002-04-10
; PRIOR APPLICATION NUMBER: 09/794,308
; PRIOR FILING DATE: 2001-02-26
; PRIOR APPLICATION NUMBER: 09/866,538
; NUMBER OF SEQ ID NOS: 78
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 238

; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-121-258-10

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-11;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
QY 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
Db 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 9

US-10-221-461-6
; Sequence 6, Application US/10221461
; Publication No. US20030092902A1
; GENERAL INFORMATION:
; APPLICANT: Marsh, Donald J.
; TITLE OF INVENTION: MELANIN CONCENTRATING HORMONE RECEPTOR
; FILE REFERENCE: 20652P
; CURRENT APPLICATION NUMBER: US/10/221.461
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: PCT/US01/08071
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/189,698
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-10-221-461-6

Query Match 97.8%; Score 1242; DB 9; Length 238;
Best Local Similarity 97.5%; Pred. No. 6.1e-11;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
Db 1 MSKGEELFTAVPLVPLVELDGDVNGHKFSVSGEGSDVYTKLTKFKICTGKLPVMPPTL 60
QY 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
Db 61 VTFPSYGVQCFSRYPDHMKRHFPEKSAPEGYVOORTIFFKDDGNYKTRAEVKFGDTLV 120
QY 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDKEDGNILGHKLENYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YGQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238
Db 181 HYQONTPIIDGPPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAGITGMDLYK 238

RESULT 10

US-09-920-922-4
; Sequence 4, Application US/09920922
; Patent No. US20020083488A1

```

; GENERAL INFORMATION:
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Sawano, Asako
; TITLE OF INVENTION: METHOD FOR MUTAGENESIS
; FILE REFERENCE: 11283-012001
; CURRENT APPLICATION NUMBER: US/09/920,922
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: JP 2000-237166
; PRIOR FILING DATE: 2000-08-04
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
US-09-920-922-4

Query Match          97.8%; Score 1242; DB 10; Length 238;
Best Local Similarity 97.5%; Pred. No. 6,1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 60
    |||||||
DB 1 MSKGEELFTGVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 60
    |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 120
    |||||||
DB 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 120
    |||||||
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
QY 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 238
    :|||||
DB 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 238
    :|||||

RESULT 11
US-09-852-000-1
; Sequence 1, Application US/09852000
; Patent No. US20020099170A1
; GENERAL INFORMATION:
; APPLICANT: Osumi, Takashi
; APPLICANT: Tsukamoto, Toshiro
; APPLICANT: Tsukamoto, No. US20020099170A1yo
; APPLICANT: Yamasaki, Masatoshi
; TITLE OF INVENTION: GREEN FLUORESCENT PROTEINS AND BLUE FLUORESCENT
; FILE REFERENCE: 046124-5005-US
; CURRENT APPLICATION NUMBER: US/09/852,000
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: JP 026418/1998
; PRIOR FILING DATE: 1998-01-23
; PRIOR APPLICATION NUMBER: US 09/121,539
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 09/615,655
; PRIOR FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; FEATURE:
; OTHER INFORMATION: Green fluorescent protein
US-09-852-000-1

Query Match          97.8%; Score 1242; DB 10; Length 238;
Best Local Similarity 97.5%; Pred. No. 6,1e-111;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 60
```

```

    |||||||
DB 1 MSKGEELFTGVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 60
    |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 120
    |||||||
DB 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 120
    |||||||
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
DB 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
QY 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 238
    :|||||
DB 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 238
    :|||||

RESULT 12
US-09-900-345A-60
; Sequence 60, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-505
; CURRENT APPLICATION NUMBER: US/09/900,345A
; PRIOR FILING DATE: 2001-07-06
; PRIOR APPLICATION NUMBER: AU P88078
; PRIOR FILING DATE: 1999-01-08
; PRIOR APPLICATION NUMBER: PCT/AU00/00008
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 185
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 60
; LENGTH: 243
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTA)5Gpp
US-09-900-345A-60

Query Match          97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1,2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 60
    :|||||
DB 6 LSKGEELFTGVVPIIVELDGDVNGHKFSVSGEGEGDVTYGKLTIKTCTTGKLPVWPPTL 65
    |||||||
QY 61 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 120
    |||||||
DB 66 VTTFSTGYOCFSRYPDHMKRHDFFKSAMPEGYVOQRTIFPKDGNKTRAEVFEGDITLV 125
    |||||||
QY 121 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
    |||||||
DB 126 NRTELKIDFKEDGNILGKLEYNNSHNVYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
    |||||||
QY 181 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 238
    :|||||
DB 186 YVOQNTPIIDGVPVLLPDNHYLSTQSALSKDPNEKRDMVLEEVTAAGITHGMDELYK 243
    :|||||

RESULT 13
US-09-900-345A-62
; Sequence 62, Application US/09900345A
; Publication No. US20030031999A1
; GENERAL INFORMATION:
; APPLICANT: Frazer, Ian Hector
; APPLICANT: Zhou, Jian
; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
; FILE REFERENCE: 10338-505
```

;; CURRENT APPLICATION NUMBER: US/09/900.345A
;; CURRENT FILING DATE: 2001-07-06
;; PRIOR APPLICATION NUMBER: AU PP8078
;; PRIOR FILING DATE: 1999-01-08
;; PRIOR APPLICATION NUMBER: PCT/AU00/00008
;; PRIOR FILING DATE: 2000-01-07
;; NUMBER OF SEQ ID NOS: 185
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 62
;; LENGTH: 243
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)5GFP
US-09-900-345A-62

Query Match 97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDDGVNGHKFSVSGEGSDVYTKLTKFKICTTGKLPVMPPTL 60
DB 6 LSKGEELFTGVVPLVLELDDGVNGHKFSVSGEGSDATYKTLTKFKICTTGKLPVMPPTL 65
QY 61 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGDNKTRAEVKEGDTLV 120
DB 66 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGDNKTRAEVKEGDTLV 125
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 243

RESULT 14
US-09-900-345A-64
;; Sequence 64, Application US/0900345A
;; Publication No. US20030031999A1
;; GENERAL INFORMATION:
;; APPLICANT: Frazer, Ian Hector
;; APPLICANT: Zhou, Jian
;; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
;; FILE REFERENCE: 10338-5US
;; CURRENT APPLICATION NUMBER: US/09/900.345A
;; CURRENT FILING DATE: 2001-07-06
;; PRIOR APPLICATION NUMBER: AU PP8078
;; PRIOR FILING DATE: 1999-01-08
;; PRIOR APPLICATION NUMBER: PCT/AU00/00008
;; PRIOR FILING DATE: 2000-01-07
;; NUMBER OF SEQ ID NOS: 185
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 64
;; LENGTH: 243
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTG)5GFP
US-09-900-345A-64

Query Match 97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDDGVNGHKFSVSGEGSDVYTKLTKFKICTTGKLPVMPPTL 60
DB 6 LSKGEELFTGVVPLVLELDDGVNGHKFSVSGEGSDATYKTLTKFKICTTGKLPVMPPTL 65
QY 61 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGDNKTRAEVKEGDTLV 120

DB 66 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGDNKTRAEVKEGDTLV 125
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 243

RESULT 15
US-09-900-345A-66
;; Sequence 66, Application US/0900345A
;; Publication No. US20030031999A1
;; GENERAL INFORMATION:
;; APPLICANT: Frazer, Ian Hector
;; APPLICANT: Zhou, Jian
;; TITLE OF INVENTION: METHOD AND POLYNUCLEOTIDES FOR DETERMINING TRANSLATIONAL
;; FILE REFERENCE: 10338-5US
;; CURRENT APPLICATION NUMBER: US/09/900.345A
;; CURRENT FILING DATE: 2001-07-06
;; PRIOR APPLICATION NUMBER: AU PP8078
;; PRIOR FILING DATE: 1999-01-08
;; PRIOR APPLICATION NUMBER: PCT/AU00/00008
;; PRIOR FILING DATE: 2000-01-07
;; NUMBER OF SEQ ID NOS: 185
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 66
;; LENGTH: 243
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Leu(CTT)5GFP
US-09-900-345A-66

Query Match 97.6%; Score 1239; DB 9; Length 243;
Best Local Similarity 97.1%; Pred. No. 1.2e-110;
Matches 231; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDDGVNGHKFSVSGEGSDVYTKLTKFKICTTGKLPVMPPTL 60
DB 6 LSKGEELFTGVVPLVLELDDGVNGHKFSVSGEGSDATYKTLTKFKICTTGKLPVMPPTL 65
QY 61 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOQRTIFFKDGDNKTRAEVKEGDTLV 120
DB 66 VTTFSSYVOCFSRYPDHMKRHDFFKSAPEGYVOERTIFFKDGDNKTRAEVKEGDTLV 125
QY 121 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
DB 126 NRTELKIDFKEDGNILGHKLEYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 185
QY 181 YYOQNPILDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
DB 186 HYQONTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 243

Search completed: July 11, 2003, 18:08:22
Job time : 54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:55:34 ; Search time 27 Seconds

(without alignments)
259.358 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGEELFAVAVPIVLVDG.....VLLEFVTAAGTTHGMDELTK 238

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents_AA:*
1: /cgn2.6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2.6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2.6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2.6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2.6/ptodata/1/1aa/PT05.COMB.pep:*
6: /cgn2.6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1270	100.0	238	3	US-08-911-825-2 Sequence 2, Appl1
2	1270	100.0	238	3	US-08-974-737-2 Sequence 2, Appl1
3	1270	100.0	238	3	US-08-706-408-2 Sequence 2, Appl1
4	1270	100.0	238	4	US-09-465-142-2 Sequence 2, Appl1
5	1246	98.1	238	1	US-08-753-143-2 Sequence 2, Appl1
6	1246	98.1	238	2	US-08-679-865-2 Sequence 2, Appl1
7	1246	98.1	238	2	US-08-680-876-2 Sequence 2, Appl1
8	1246	98.1	238	2	US-08-792-553-2 Sequence 2, Appl1
9	1246	98.1	238	3	US-08-753-144-2 Sequence 2, Appl1
10	1246	98.1	238	4	US-09-094-359-2 Sequence 2, Appl1
11	1246	98.1	238	4	US-09-172-063-2 Sequence 2, Appl1
12	1246	98.1	238	4	US-09-263-975-2 Sequence 2, Appl1
13	1246	98.1	238	4	US-08-727-452-2 Sequence 2, Appl1
14	1246	98.1	238	4	US-08-418-785-1 Sequence 2, Appl1
15	1242	97.8	238	1	US-08-337-915A-2 Sequence 1, Appl1
16	1242	97.8	238	4	US-09-121-539-1 Sequence 1, Appl1
17	1242	97.8	238	5	PCR-US95-14692-2 Sequence 2, Appl1
18	1241	97.7	239	3	US-08-646-538-2 Sequence 2, Appl1
19	1241	97.7	239	3	US-09-503-222-2 Sequence 2, Appl1
20	1239	97.6	1070	4	US-09-091-042A-2 Sequence 2, Appl1
21	1239	97.6	1452	4	US-09-127-227-2 Sequence 2, Appl1
22	1233	97.1	238	3	US-08-893-327-16 Sequence 16, Appl
23	1233	97.1	247	3	US-08-893-327-18 Sequence 18, Appl
24	1233	97.1	247	3	US-08-893-327-20 Sequence 20, Appl
25	1232	97.0	238	4	US-09-213-343-4 Sequence 4, Appl1
26	1230	96.9	238	2	US-08-588-201-2 Sequence 2, Appl1
27	1230	96.9	238	2	US-09-169-605-2 Sequence 2, Appl1

28	1230	96.9	238	3	US-08-893-327-2	Sequence 2, Appl1
29	1226	96.5	238	3	US-08-643-704A-49	Sequence 49, Appl
30	1224	96.4	239	4	US-09-121-539-14	Sequence 14, Appl
31	1221	96.1	238	1	US-08-452-295-1	Sequence 1, Appl1
32	1221	96.1	238	4	US-09-172-063-8	Sequence 8, Appl1
33	1221	96.1	238	4	US-09-316-919-9	Sequence 9, Appl1
34	1220	96.1	239	4	US-09-513-783A-52	Sequence 52, Appl
35	1220	96.1	282	4	US-09-513-783A-14	Sequence 14, Appl
36	1220	96.1	287	4	US-09-513-783A-8	Sequence 8, Appl1
37	1220	96.1	291	4	US-09-513-783A-10	Sequence 10, Appl
38	1220	96.1	292	4	US-09-513-783A-16	Sequence 16, Appl
39	1220	96.1	295	4	US-09-513-783A-20	Sequence 20, Appl
40	1220	96.1	296	4	US-09-513-783A-12	Sequence 12, Appl
41	1220	96.1	302	4	US-09-513-783A-26	Sequence 26, Appl
42	1220	96.1	316	4	US-09-513-783A-24	Sequence 24, Appl
43	1220	96.1	326	4	US-09-513-783A-28	Sequence 28, Appl
44	1220	96.1	347	4	US-09-513-783A-30	Sequence 30, Appl
45	1220	96.1	350	4	US-09-513-783A-30	Sequence 30, Appl

ALIGNMENTS

```
RESULT 1
US-08-911-825-2
; Sequence 2, Application US/08911825
; Patent No. 6054321
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/911,825
; FILING DATE: 15-AUG-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-911-825-2

Query Match      100.0%  Score 1270;  DB 3;  Length 238;
Best Local Similarity 100.0%;  Pred. No. 1.5e-130;
Matches 238;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

OY      1 MSKGEELFAVAVPIVLVDGNGHKEFSVSGEGEDVYTGKTLTFTGKLPVWPPL 60
      |||
DB      1 MSKGEELFAVAVPIVLVDGNGHKEFSVSGEGEDVYTGKTLTFTGKLPVWPPL 60
      |||
OY      61 VTFYSYGVCCFSRYPDHMKRHDFKSAPEGVVQORTIFFKDDGNYKTRAEVKFEEDTLV 120
      |||
```

Db 61 VTTSYGVQCFSRPDMKRRDFEKSAMPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 2

US-08-974-737-2
; Sequence 2, Application US/08974737
; Patent No. 6077707
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,737
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/911,825
; FILING DATE: 15-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/056001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-974-737-2

Query Match 100.0%; Score 1270; DB 3; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVILVELDGDVNGHKFSVSGEGDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVILVELDGDVNGHKFSVSGEGDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTSYGVQCFSRPDMKRRDFEKSAMPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
Db 61 VTTSYGVQCFSRPDMKRRDFEKSAMPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 3

US-08-706-408-2
; Sequence 2, Application US/08706408
; Patent No. 6124128
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/706,408
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/024,050
; FILING DATE: 16-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Halle, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/029001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-706-408-2

Query Match 100.0%; Score 1270; DB 3; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.5e-130;
Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVILVELDGDVNGHKFSVSGEGDVTYGLTKLFICTTGKLPVMPPTL 60
Db 1 MSKGEELFTAVVPLVILVELDGDVNGHKFSVSGEGDVTYGLTKLFICTTGKLPVMPPTL 60
QY 61 VTTSYGVQCFSRPDMKRRDFEKSAMPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
Db 61 VTTSYGVQCFSRPDMKRRDFEKSAMPEGYVOQRTIFFKDDGNKTRAEVKEGDTLV 120
QY 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
Db 121 NRIELKIDFKEDGNILGHKLENYNSHNHYIMADKOKNGIKVFKIRHNIEDGSSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238
Db 181 YVOQNTPIIDGPVLLPDMHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITHGMDLYK 238

RESULT 4

US-09-465-142-2
; Sequence 2, Application US/09465142
; Patent No. 6403374
; GENERAL INFORMATION:
; APPLICANT: Tsien et al., Roger Y. et al.
; TITLE OF INVENTION: LONG WAVELENGTH MUTANT FLUORESCENT

TITLE OF INVENTION: PROTEINS
 NUMBER OF SEQUENCES: 4
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: CA
 COUNTRY: USA
 ZIP: 92037
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/465,142
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/974,737
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 07257/056001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-465-142-2

Query Match 100.0%; Score 1270; DB 4; Length 238;
 Best Local Similarity 100.0%; Pred. No. 1.5e-130;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 DB 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEKSAMPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
 DB 61 VTFESYGVQCFSRYPDHMKRHDFEKSAMPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKVNFKIRINIEDGSVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKVNFKIRINIEDGSVQLAD 180
 QY 181 YVQNTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238
 DB 181 YVQNTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238

RESULT 5
 US-08-753-143-2
 Sequence 2, Application US/08753143A
 GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Heim, Roger
 TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
 FILE REFERENCE: 07257/032003
 CURRENT APPLICATION NUMBER: US/08/753,143A
 CURRENT FILING DATE: 1996-11-20
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 2
 LENGTH: 238
 TYPE: PRT
 ORGANISM: Aequorea victoria

US-08-753-143-2

Query Match 98.1%; Score 1246; DB 1; Length 238;
 Best Local Similarity 97.9%; Pred. No. 6.1e-128;
 Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 DB 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEKSAMPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
 DB 61 VTFESYGVQCFSRYPDHMKRHDFEKSAMPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKVNFKIRINIEDGSVQLAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKOKNGIKVNFKIRINIEDGSVQLAD 180
 QY 181 YVQNTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238
 DB 181 YVQNTPIIDGFPVLLPDNHYLSTQSALSKDPNEKRDHMYLLEFVTAAGITHGMDELTK 238

RESULT 6
 US-08-679-865-2
 Sequence 2, Application US/08679865
 Patent No. 5912137

GENERAL INFORMATION:
 APPLICANT: Tsien, Roger Y.
 APPLICANT: Cubitt, Andrew B.
 TITLE OF INVENTION: Assays for Protein Kinases Using
 TITLE OF INVENTION: Fluorescent Protein Substrates
 NUMBER OF SEQUENCES: 48
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Townsend and Crew LLP
 STREET: Two Embarcadero Center, Eighth Floor
 CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/679,865
 FILING DATE: 16-JUL-1996
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
 NAME: Storella, John S.
 REGISTRATION NUMBER: 32,944
 REFERENCE/DOCKET NUMBER: 023072-069000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 238 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-679-865-2

Query Match 98.1%; Score 1246; DB 2; Length 238;
 Best Local Similarity 97.9%; Pred. No. 6.1e-128;
 Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 DB 1 MSKGEELTAVVPIILVELDGVNGHKFSVSGEGEDVYTGKLTLEFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEKSAMPEGVVQORTIFFKDDGNVKTAEVKEFGDITLV 120

|||||
Db 61 VTTFSYGVQCFSRYPDHMKRHRDFPKSAMPEGYVQERTIFFEKDDGNYKTRAEVKREGDTLV 120
QY 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
QY 181 YVQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238
Db 181 HYQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238

RESULT 7

US-08-680-876-2
; Sequence 2, Application US/08680876
; Patent No. 5925558
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Cubitt, Andrew B.
; TITLE OF INVENTION: Assays for Protein Kinases Using
; TITLE OF INVENTION: Fluorescent Protein Substrates
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680, 876
; FILING DATE: 16-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Storella, John S.
; REGISTRATION NUMBER: 32,944
; REFERENCE/DOCKET NUMBER: 02307Z-069200
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-680-876-2

Query Match 98.1%; Score 1246; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVPLVELDGVNKGKFSVSGEGSDVYTGKLTLEKFICTTGKLPVPVPTL 60
Db 1 MSKGEELFTGVVPLVELDGVNKGKFSVSGEGSDVYTGKLTLEKFICTTGKLPVPVPTL 60
QY 61 VTTFSYGVQCFSRYPDHMKRHRDFPKSAMPEGYVQERTIFFEKDDGNYKTRAEVKREGDTLV 120
Db 61 VTTFSYGVQCFSRYPDHMKRHRDFPKSAMPEGYVQERTIFFEKDDGNYKTRAEVKREGDTLV 120
QY 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
QY 181 YVQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238
Db 181 HYQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238

RESULT 8

US-08-792-553-2
; Sequence 2, Application US/08792553
; Patent No. 5981200
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Helm, Roger
; TITLE OF INVENTION: Tandem Fluorescent Protein Constructs
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & RICHARDSON P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/792,553
; FILING DATE: 31-JAN-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lisa A. Halle, Ph.D.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/041001/UC 96-160-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-792-553-2

Query Match 98.1%; Score 1246; DB 2; Length 238;
Best Local Similarity 97.9%; Pred. No. 6.1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVPLVELDGVNKGKFSVSGEGSDVYTGKLTLEKFICTTGKLPVPVPTL 60
Db 1 MSKGEELFTGVVPLVELDGVNKGKFSVSGEGSDVYTGKLTLEKFICTTGKLPVPVPTL 60
QY 61 VTTFSYGVQCFSRYPDHMKRHRDFPKSAMPEGYVQERTIFFEKDDGNYKTRAEVKREGDTLV 120
Db 61 VTTFSYGVQCFSRYPDHMKRHRDFPKSAMPEGYVQERTIFFEKDDGNYKTRAEVKREGDTLV 120
QY 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
Db 121 NRIELKGIIDFKEDGNILGHKLEIYNNSHNYIMADKONGIKVFKIRHNIEDSSVOLAD 180
QY 181 YVQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238
Db 181 HYQONTPIIDGVPVLLPDNHYLSTQSALSKEPDNEKRDMHVLLEFYTAAGITGHGMDLYK 238

RESULT 9

US-08-753-144-2
; Sequence 2, Application US/08753144
; Patent No. 6066476
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Helm, Roger
; TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.

```

; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,144
; FILING DATE: 20-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/727,452
; FILING DATE: 10-OCT-1996
; APPLICATION NUMBER: US95/14692
; FILING DATE: 13-NOV-1995
; APPLICATION NUMBER: 08/337,915
; FILING DATE: 10-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07257/032002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-08-753-144-2

Query Match      98.1%; Score 1246; DB 3; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVNGHKFSVSGEGSDVYTGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYTGKLTFTCTTGKLPVPMPTL 60

QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

QY 181 YVQONTPLDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238
DB 181 HYQONTPLIGDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238

RESULT 10
US-09-094-359-2
; Sequence 2, Application US/09094359
; Patent No. 6140132
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/067001
; CURRENT APPLICATION NUMBER: US/09/094,359
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

```

```

; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-094-359-2

Query Match      98.1%; Score 1246; DB 4; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVNGHKFSVSGEGSDVYTGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYTGKLTFTCTTGKLPVPMPTL 60

QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

QY 181 YVQONTPLDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238
DB 181 HYQONTPLIGDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238

RESULT 11
US-09-172-063-2
; Sequence 2, Application US/09172063
; Patent No. 6150176
; GENERAL INFORMATION:
; APPLICANT: Tsien, Roger Y.
; APPLICANT: Miyawaki, Atsushi
; APPLICANT: Llopis, Juan
; APPLICANT: Wachter, Rebekka M.
; APPLICANT: Remington, S. James
; TITLE OF INVENTION: FLUORESCENT PROTEIN SENSORS FOR
; FILE REFERENCE: 07257/071001
; CURRENT APPLICATION NUMBER: US/09/172,063
; EARLIER FILING DATE: 1998-10-13
; EARLIER FILING DATE: 1998-06-09
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Aequorea victoria
; US-09-172-063-2

Query Match      98.1%; Score 1246; DB 4; Length 238;
Best Local Similarity 97.9%; Pred. No. 6,1e-128;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVNGHKFSVSGEGSDVYTGKLTFTCTTGKLPVPMPTL 60
DB 1 MSKGEELFTGVVPLVLELDGVNGHKFSVSGEGSDATYTGKLTFTCTTGKLPVPMPTL 60

QY 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120
DB 61 VTTFSYGVQCFSRYPDHKKRHDFFKSAPEGYVOERTTFEKDDGNYKTRAEVKEGDTLV 120

QY 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDNIIHGKLEYNVNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

QY 181 YVQONTPLDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238
DB 181 HYQONTPLIGDGPVLLPDNHYLSTOSALSCKDPNEKRDMHWLLEFYTAAGITGMDLYK 238

RESULT 12

```

```

CURRENT APPLICATION NUMBER: US/08/727,452A
CURRENT FILING DATE: 1996-03-20
EARLIER APPLICATION NUMBER: PCT/US95/14692
EARLIER FILING DATE: 1995-11-13
EARLIER APPLICATION NUMBER: US 07/337,915
EARLIER FILING DATE: 1994-11-10
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria

Query Match
Best Local Similarity 98.1%; Score 1246; DB 4; Length 238;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

1 MSKGEELFTAVVPIVLVLELDGDVNGHKFSVSGEGEGDVTYKLLKFLCTTGKLPVWPPTL 60
|||||
1 MSKGEELFTGVVPIVLVLELDGDVNGHKFSVSGEGEGDVTYKLLKFLCTTGKLPVWPPTL 60

Db
61 VTTFSSVVOCFSRPDHMKRHHDFKSSAMPESGYOQRIFFPKDDGNTKTRAEVKPEGGTLV 120
|||||
61 VTTFSSVVOGCSRPDMMKRRHDFKSSAMPESGYOQRIFFPKDDGNTKTRAEVKPEGGTLV 120

Qy
121 NRLELKIDFKEDGNILGHRLELYNYSNHNYIMADKOKNOIKVFKIRHNIEDGSYQAD 180
|||||
121 NRLELKIDFKEDGNILGHRLELYNYSNHNYIMADKOKNGIKVFKIRHNIEDGSYQAD 180

Db
181 YVQONTPLDGPVLLPNNHLSSTOSALSCKDPNEKRDMVLLFPTAAGITHGMDLYK 238
|||||
181 HYQONTPIGDGPVLLPNNHLSSTOSALSCKDPNEKRDMVLLFPTAAGITHGMDLYK 238

RESULT 14
US-09-418-785-1
Sequence 1, Application US/09418785
Patent No. 6414119
GENERAL INFORMATION:
APPLICANT: Fisher, Hugh
TITLE OF INVENTION: Rapidly Greening, Low Oxygen Mutant of
FILE REFERENCE: RUTC 99-0011
CURRENT APPLICATION NUMBER: US/09/418,785
CURRENT FILING DATE: 1999-10-15
PRIOR APPLICATION NUMBER: 60/104,563
PRIOR FILING DATE: 1998-10-16
NUMBER OF SEQ ID NOS: 3
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 238
TYPE: PRT
ORGANISM: Aequorea victoria
PUBLICATION INFORMATION:
AUTHORS: Prasher, D.C. et al.
TITLE: Primary structure of the Aequorea victoria green-f
JOURNAL: Gene
VOLUME: 111
PAGES: 229-233
DATE: 1992-01-01
DATABASE ACCESSION NUMBER: Genbank M62653
DATABASE ENTRY DATE: 1993-04-26
US-09-418-785-1

Query Match
Best Local Similarity 98.1%; Score 1246; DB 4; Length 238;
Matches 233; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

1 MSKGEELFTAVVPIVLVLELDGDVNGHKFSVSGEGEGDVTYKLLKFLCTTGKLPVWPPTL 60
|||||
1 MSKGEELFTGVVPIVLVLELDGDVNGHKFSVSGEGEGDVTYKLLKFLCTTGKLPVWPPTL 60

```

QY 61 VTTFSGVQCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFPKDDGNKYTRAEVKFEQDTLV 120
|||||
Db 61 VTTFSGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNKYTRAEVKFEQDTLV 120
|||||
QY 121 NRLEKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
|||||
Db 121 NRLEKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
|||||
QY 181 YVOQNTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITGHMDELK 238
:|||||
Db 181 HYQONTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITGHMDELK 238
:|||||

RESULT 15

US-08-337-915A-2
: Sequence 2, Application US/08337915A
: Patent No. 5625048
: GENERAL INFORMATION:
: APPLICANT: Tsien, Roger Y.
: APPLICANT: Heim, Roger
: TITLE OF INVENTION: MODIFIED GREEN FLUORESCENT PROTEINS
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Robbins, Berliner & Carson
: STREET: 201 No. 5625048th Figueroa Street, Suite 500
: CITY: Los Angeles
: STATE: California
: COUNTRY: USA
: ZIP: 90012
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentln Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/337,915A
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spitals, John P.
: REGISTRATION NUMBER: 29,215
: REFERENCE/DOCKET NUMBER: 1279-178
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (213) 977-1001
: TELEFAX: (213) 977-1003
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 238 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-337-915A-2

Query Match 97.8%; Score 1242; DB 1; Length 238;
Best Local Similarity 97.5%; Pred. No. 1.7e-127;
Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLVLELDGVDNGHKSFSVSGEGDVTYGGKLTLEKICTTGKLPVPWPTL 60
|||||
Db 1 MSKGEELFTGVVPLVLELDGVDNGHKSFSVSGEGDVTYGGKLTLEKICTTGKLPVPWPTL 60
|||||
QY 61 VTTFSGVQCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNKYTRAEVKFEQDTLV 120
|||||
Db 61 VTTFSGVQCFSRYPDHKKRHDFFKSAMPEGYVOERTIFFKDDGNKYTRAEVKFEQDTLV 120
|||||
QY 121 NRLEKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
|||||
Db 121 NRLEKGIIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNFKIRHNIEDGSVOLAD 180
|||||
QY 181 YVOQNTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITGHMDELK 238
:|||||
Db 181 HYQONTPIIDGPPVLLPDNHYLSTOSALSKDPEKRDHMYLLEFVTAAGITGHMDELK 238
:|||||

Search completed: July 11, 2003, 18:00:06
Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:51:43 ; Search time 81 Seconds
(Without alignments)
605.422 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGEELFTAVVPLVELDQ.....VLEFVTAAGITGMDELYK 238

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.8	238	5	Q27903 unidentified
2	1235	97.2	238	5	Q93125 aequorea vi
3	1200	94.5	238	5	Q17105 aequorea vi
4	1185	93.3	238	5	Q17106 aequorea vi
5	1182	93.1	758	4	Q96J01 homo sapien
6	1079	85.0	238	5	Q8WP95 aequorea ma
7	1073	84.5	238	5	Q8WTC6 aequorea ma
8	1065	83.9	238	5	Q8WTC4 aequorea ma
9	1063	83.7	238	5	Q8WTC0 aequorea ma
10	1062	83.6	238	5	Q8WTC9 aequorea ma
11	1062	83.6	238	5	Q8WTC8 aequorea ma
12	1060	83.5	238	5	Q8WTC7 aequorea ma
13	1058	83.3	238	5	Q8WTC5 aequorea ma
14	255.5	20.1	225	5	Q95UAT montastrea
15	245	19.3	225	5	Q963F5 montastrea
16	241.5	19.0	266	5	Q9U6Y3 clavularia

17	239.5	18.9	225	5	Q8T5F1 montastrea
18	239.5	18.9	236	5	Q8T6D0 dendroneph
19	212.5	16.7	235	5	Q8T5R0 scolymia cu
20	212.5	16.7	238	5	Q9BLT9 renilla mu
21	209.5	16.5	225	5	Q9U6Y8 discosoma s
22	209	16.5	229	5	Q9U6Y6 anemoma ma
23	206.5	16.3	234	5	Q8T5F2 montastrea
24	205	16.1	232	5	Q9U6Y7 discosoma s
25	204	16.1	227	5	Q962E9 montastrea
26	203	16.0	221	5	Q95P04 gonioporta t
27	200.5	15.8	232	5	Q9GPI5 anemoma su
28	200	15.7	227	5	Q95VW0 montastrea
29	199.5	15.7	225	5	Q8T6P9 heteractis
30	196.5	15.5	232	5	Q9GZT8 heteractis
31	196	15.4	230	5	Q9GCT7 discosoma s
32	191.5	15.1	233	5	Q963I9 renilla ren
33	190.5	15.0	238	5	Q9BLT0 pillosarcus
34	189.5	14.9	228	5	Q9GPI6 anemoma su
35	183.5	14.4	227	5	Q95W85 heteractis
36	183.5	14.4	231	5	Q9U6Y5 zoanthus sp
37	176.5	13.9	229	5	Q8T5R7 condylactis
38	175.5	13.8	227	5	Q95W86 condylactis
39	173	13.6	231	5	Q9U6Y4 zoanthus sp
40	169.5	13.3	227	5	Q95W11 condylactis
41	168.5	13.3	231	5	Q8T4D4 zoanthus sp
42	168	13.2	234	5	Q8T5F3 scolymia cu
43	165	13.0	231	5	Q8T5F8 ricordea fl
44	159.5	12.6	231	5	Q8T6T8 discosoma s
45	151	11.9	231	5	Q8T5E9 ricordea fl

ALIGNMENTS

RESULT 1

ID Q27903 PRELIMINARY: PRT; 238 AA.
AC Q27903:
DT 01-NOV-1996 (TREMBLrel. 01. Created)
DT 01-NOV-1996 (TREMBLrel. 01. Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21. Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS unidentified.
OC unclassified.
OX NCBI_TaxID=32644;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97299832; PubMed=9154981;
RA Rouwendal G.J.A., Mendes O., Wolbert E.J.H., de Boer A.D.;
RT "Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage."
RL Plant Mol. Biol. 33:989-999(1997).
DR EMBL; X96418; CAA65278.1; -
DR HSSP; P42212; IGFL.
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP.1.
DR PRINTS; PR01229; GFPDORESCENT.
DR PRODom; PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 26884 MW; CA932D47262AF2D3 CRC64;

Query Match Best Local Similarity 97.8%; Pred. No. 7.7e-94; Length 238;

Matches 232; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY	1	MSKGEELFTAVVPLVELDQVNGHKFSVSGEGSDVYGGKLTLLKFTCTGKLPVMPPTL	60
DB	1	MSKGEELFTAVVPLVELDQVNGHKFSVSGEGSDVYGGKLTLLKFTCTGKLPVMPPTL	60
QY	61	VTTSTGYQCFSRYPDHHKRRDFRSAPMEGYVOORTFFFDGNGYKRAVKEGDTLV	120
DB	61	VTTSTGYQCFSRYPDHHKRRDFRSAPMEGYVOORTFFFDGNGYKRAVKEGDTLV	120

QY 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 Db 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONFTPLDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 Db 181 HYQONTPIGDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 2

ID 093125 PRELIMINARY; PRT: 238 AA.
 AC 093125;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein mutant 3.
 GN GFP.
 OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96305137; PubMed=8707053;
 RA Cormack B.P., Valdivia R.H., Falkow S.;
 RT "FACS-Optimized mutants of the green fluorescent protein (GFP).";
 RL Gene 173:33-38(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Cormack B.P., Egerton M., Gow N.A.R., Falkow S.,
 RA Brown A.J.P.;
 RT "Yeast Enhanced Green Fluorescent Protein (yEGFP): a reporter of gene
 expression in Candida albicans";
 RL Microbiology 0:0-0(1996).
 DR EMBL: U73901; AAB18957.1; -
 DR HSSP: P42212; 1BFP.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 SQ SEQUENCE 238 AA; 26840 MW; A28622809A3DEA60 CRC64;

Query Match 97.2%; Score 1235; DB 5; Length 238;
 Best Local Similarity 96.6%; Pred. No. 2.9e-93;

Matches 230; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 Db 1 MSKGEELFTGVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFPKDDGNKTRAEVFEGDTLV 120
 Db 61 VTFESYGVQCFARYPDHMKRHDFEFKSAMPEGYVQERTIFPKDDGNKTRAEVFEGDTLV 120
 QY 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 Db 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONFTPLDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 Db 181 HYQONTPIGDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 3

ID 017105 PRELIMINARY; PRT: 238 AA.
 AC 017105;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein (Fragment).
 GN GFP.

OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watkins J.N., Campbell A.K.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X83959; CAA58789.1; -
 DR HSSP: P42212; 1GFL.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 FT NON_TER 238
 SQ SEQUENCE 238 AA; 26950 MW; 26E2BE450E748E44 CRC64;

Query Match 94.5%; Score 1200; DB 5; Length 238;
 Best Local Similarity 93.3%; Pred. No. 2.1e-90;

Matches 222; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 Db 1 MSKGEELFTGVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFPKDDGNKTRAEVFEGDTLV 120
 Db 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQERTIFPKDDGNKTRAEVFEGDTLV 120
 QY 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 Db 121 NRLEKIDKREKDNIGLHLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOONFTPLDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238
 Db 181 HYQONTPIGDGPVLLPNDNHYLSTOSALSKDPNEKRDMHVLLEFVTAAGITHGMDLYK 238

RESULT 4

ID 017106 PRELIMINARY; PRT: 238 AA.
 AC 017106;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Green fluorescent protein (Fragment).
 GN GFP.
 OS Aequorea victoria (Jellyfish).
 OC Eukaryota; Metazoa; Chnidaria; Hydrozoa; Hydroida; Leptomedusae;
 OC Aequoreidae; Aequorea.
 OX NCBI_TaxID=6100;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watkins J.N., Campbell A.K.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X83960; CAA58790.1; -
 DR HSSP: P42212; 1BFP.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PR01229; GFPLORESCENT.
 DR PRODOM: PD013756; Green_fl_protein; 1.
 FT NON_TER 238
 SQ SEQUENCE 238 AA; 26867 MW; BD4648262D8EABD4 CRC64;

Query Match 93.3%; Score 1185; DB 5; Length 238;
 Best Local Similarity 92.4%; Pred. No. 3.5e-89;

Matches 220; Conservative 9; Mismatches 9; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 Db 1 MSKGEELFTGVVPLIVELDGDVNGHKFSVSGEGSDVTYGLTKLTKFICTTGKLPVWPPTL 60
 QY 61 VTFESYGVQCFSRYPDHMKRHDFEFKSAMPEGYVQARTIFPKDDGNKTRAEVFEGDTLV 120


```

Db      61 VTFESYGVOCFSRPDHNKOHDFLKSAMEEGYVQERTIFYKDDGNYKTRAEEKFEGDITLV 120
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMAKONGKIVNKRIRHNIEDSGVOLAD 180
        |||
Db      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMGKPNKIKVNRKIRHNIEDSGVOLAD 180
        |||
QY      181 YVQONTPLIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGHMDELX 238
        |||
Db      181 HYQONTPLIGDGFVLLPDNHYISTOSALSQDPNKRHDHMYLLEFVTAAGITGHMDELX 238
        |||

```

RESULT 5

```

096J01  PRELIMINARY: PRT: 758 AA.
ID      096J01
AC      096J01:
DT      01-DEC-2001 (TREMBLrel. 19, Created)
DT      01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Ralchu404X.
GN      RALCHU404X.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RC      SEQUENCE FROM N.A.
RT      TISSUE=SPLEEN.
RX      MEDLINE=21322811; PubMed=11429608;
RA      Mochizuki N., Yamashita S., Kurokawa K., Ohba Y., Nagai T.,
RA      Miyawaki A., Matsuda M.;
RT      "Spacio-temporal Images of Growth Factor-Induced Activation of Ras and
RT      Rap1.";
RL      Nature 411:1065-1068(2001).
DR      EMBL; AB051846; BAB61868.1;
DR      InterPro; IPR000786; Green_fl_protein.
DR      InterPro; IPR001230; Prenyl_site.
DR      InterPro; IPR001806; Ras_trnsfmrng.
DR      InterPro; IPR003116; RBD.
DR      InterPro; IPR005225; Small_GTP.
DR      Pfam; PF01353; GFP; 2.
DR      Pfam; PF00071; ras; 1.
DR      Pfam; PF02196; RBD; 1.
DR      ProDom; PD013756; Green_fl_protein; 2.
DR      TrIRFAMS; TIGR00231; small_GTP; 1.
DR      PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
KW      GTP-binding. 758 AA; 85015 MW; 8612408F607CFD49 CRC64;
SQ      SEQUENCE

```

Query Match 93.1%; Score 1182; DB 4; Length 758;
 Best Local Similarity 94.1%; Pred. No. 2.7e-88;
 Matches 222; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

```

QY      1 MSKGEELTAVVPIVLLEDGVDNGHGFVSFGEGEDVYTGKTLFICTGKLPVWPPTL 60
        |||
Db      499 VSKGEELTGVVPIVLLEDGVDNGHGFVSFGEGEDATYKTLFICTGKLPVWPPTL 558
        |||
QY      61 VTFESYGVOCFSRPDHNKOHDFLKSAMEEGYVQERTIFYKDDGNYKTRAEEKFEGDITLV 120
        |||
Db      559 VTTLTWGVOCFSRPDHNKOHDFLKSAMEEGYVQERTIFYKDDGNYKTRAEEKFEGDITLV 618
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNRKIRHNIEDSGVOLAD 180
        |||
Db      619 NRTELKIDFKEDGNILGHKLEYNNSHNHYITADKOKNGIKVNRKIRHNIEDSGVOLAD 678
        |||
QY      181 YVQONTPLIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGHMDEL 236
        |||
Db      679 HYQONTPLIGDGFVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGHMDEL 734
        |||

```

RESULT 6

```

08WP95  PRELIMINARY: PRT: 238 AA.
ID      08WP95

```

```

AC      08WP95;
DT      01-MAR-2002 (TREMBLrel. 20, Created)
DT      01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Green fluorescent protein.
GN      GFPX.
OS      Aequorea macrodactyla.
OC      Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC      Aequoreidae; Aequorea.
OX      NCBI_TaxID=147615;
RN      [1]
RC      SEQUENCE FROM N.A.
RP      STRAIN=GFPX, AND GFPDNAMEX;
RA      Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Qin Y.X., Pang S.O.,
RA      Li S.J., Xia N.S.;
RL      Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AY013824; AAK02052.1;
DR      EMBL; AY013821; AAK02059.1;
DR      InterPro; IPR000786; Green_fl_protein.
DR      Pfam; PF01353; GFP; 1.
DR      ProDom; PD013756; Green_fl_protein; 1.
SQ      SEQUENCE 238 AA; 27049 MW; 8185D0E5E529012B CRC64;

```

Query Match 85.0%; Score 1079; DB 5; Length 238;
 Best Local Similarity 81.5%; Pred. No. 1.6e-80;
 Matches 194; Conservative 21; Mismatches 23; Indels 0; Gaps 0;

```

QY      1 MSKGEELTAVVPIVLLEDGVDNGHGFVSFGEGEDVYTGKTLFICTGKLPVWPPTL 60
        |||
Db      1 MSKGEELTGVVPIVLLEDGVDNGHGFVSFGEGEDATYKTLFICTGKLPVWPPTL 60
        |||
QY      61 VTFESYGVOCFSRPDHNKOHDFLKSAMEEGYVQERTIFYKDDGNYKTRAEEKFEGDITLV 120
        |||
Db      61 VTFESYGVOCFSRPDHNKOHDFLKSAMEEGYVQERTIFYKDDGNYKTRAEEKFEGDITLV 120
        |||
QY      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYIMADKOKNGIKVNRKIRHNIEDSGVOLAD 180
        |||
Db      121 NRTELKIDFKEDGNILGHKLEYNNSHNHYITADKOKNGIKVNRKIRHNIEDSGVOLAD 180
        |||
QY      181 YVQONTPLIDGFPVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGHMDELX 238
        |||
Db      181 HYQONTPLIGDGFVLLPDNHYISTOSALSQDPNEKRDHMYLLEFVTAAGITGHMDELX 238
        |||

```

RESULT 7

```

ID      08WTC6  PRELIMINARY: PRT: 238 AA.
AC      08WTC6;
DT      01-MAR-2002 (TREMBLrel. 20, Created)
DT      01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Green fluorescent protein.
GN      GFP.
OS      Aequorea macrodactyla.
OC      Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC      Aequoreidae; Aequorea.
OX      NCBI_TaxID=147615;
RN      [1]
RC      SEQUENCE FROM N.A.
RP      STRAIN=GFPX190V;
RA      Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.O., Chen M.,
RA      Li S.J., Xia N.S.;
RT      "Colorful mutants of green fluorescent protein from Aequorea
RT      macrodactyla.";
RL      Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR      EMBL; AF435431; AAL33916.1;
DR      InterPro; IPR000786; Green_fl_protein.
DR      Pfam; PF01353; GFP; 1.
DR      ProDom; PD013756; Green_fl_protein; 1.
SQ      SEQUENCE 238 AA; 27015 MW; 6B88FD75E88926903 CRC64;

```

Query Match 84.5%; Score 1073; DB 5; Length 238;
 Best Local Similarity 81.1%; Pred. No. 5e-80;


```

QY 1 MSKGEELFTAVVPIVLVDGNGHKSFSVSGEGDVTYTKLTIKFICTTGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRPRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTIGYGIQCFARYPEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGKVEGDTLV 120
QY 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITGMDLYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQNVNPLGDGPVLLPINHYLSFQTALSKDRNETRDHNVFLFEFSAGCHTGMDELTK 238

```

RESULT 11

```

Q8WTC8 PRELIMINARY: PRT: 238 AA.
ID 08WTC8
AC 08WTC8:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM163;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF435429; AAL33914.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27047 MW; 5F80A18FA1E7C84D CRC64;

```

Query Match 83.6%; Score 1062; DB 5; Length 238;
 Best Local Similarity 80.3%; Pred. No. 4e-79;
 Matches 191; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGNGHKSFSVSGEGDVTYTKLTIKFICTTGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRPRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTIGYGIQCFARYPEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGKVEGDTLV 120
QY 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITGMDLYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQNVNPLGDGPVLLPINHYLSYQTALSKDRNETRDHNVFLFEFSAGCHTGMDELTK 238

```

RESULT 12

```

Q8WTC7 PRELIMINARY: PRT: 238 AA.
ID 08WTC7
AC 08WTC7:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Green fluorescent protein.

```

```

GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GFPM191U;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF435430; AAL33915.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27002 MW; BD5BA2982264C018 CRC64;

```

Query Match 83.5%; Score 1060; DB 5; Length 238;
 Best Local Similarity 80.3%; Pred. No. 5.8e-79;
 Matches 191; Conservative 21; Mismatches 26; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGNGHKSFSVSGEGDVTYTKLTIKFICTTGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 MSKGEELFTGIIVPLIELDGDVHGKFSVRGEGDADYKLEIKFICTTGKLPVWPPTL 60
QY 61 VTTFSYGVQCSRPRPDHKKRDFKRSAMPEGYVOORTFFDDGNGYKTRAAYKEGDTLV 120
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 VTTIGYGIQCFARYPEHKMDFKRSAMPEGYIOERTFFDDGKYKTRGKVEGDTLV 120
QY 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 NRIELKGIDPEKEDNIIGHKLEYNNSHNVTIMADKONGIKVNFKIRHNIEDSGVOLAD 180
QY 181 YYQONTPLDGPVLLPDNHYLSTQSALSKDPNEKRDMHVLLEFYTAAGITGMDLYK 238
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 HYQNVNPLGDGPVLLPINHYLSYQTALSKDRNETRDHNVFLFEFSAGCHTGMDELTK 238

```

RESULT 13

```

Q8WTC5 PRELIMINARY: PRT: 238 AA.
ID 08WTC5
AC 08WTC5:
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Orange fluorescent protein.
GN GFP.
OS Aequorea macrodactyla.
OC Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroida; Leptomedusae;
OC Aequoreidae; Aequorea.
CX NCBI_TaxID=147615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OPPM;
RA Luo W.X., Zhang J., Yang H.J., Xie X.Y., Li S.W., Guan B.Q., Chen M.,
RA Li S.J., Xia N.S.;
RT "Colorful mutants of green fluorescent protein from Aequorea
RT macrodactyla."
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF435432; AAL33917.1;
DR InterPro: IPR000786; Green_fl_protein.
DR Pfam: PF01353; GFP; 1.
DR ProDom: PD013756; Green_fl_protein; 1.
SQ SEQUENCE 238 AA; 27018 MW; 75521BA5534E573A CRC64;

```

Query Match 83.3%; Score 1058; DB 5; Length 238;
 Best Local Similarity 81.1%; Pred. No. 8.5e-79;
 Matches 193; Conservative 18; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 MSKGEELFTAVVPIVLVDGNGHKSFSVSGEGDVTYTKLTIKFICTTGKLPVWPPTL 60
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

DB 1 MSKGEELGTGVPVLVELDGVNGHKFSVSGEGSDVYTKLTKFICTTGKLPVMPPTL 60
 QY 61 VTFPSYGVOCFSRYPDHMKRHFESKAMPBGVYQORTIFFKDDGNYKTRAEVKEGGDTLV 120
 DB 61 VTLTGVIILCFARTRPEMKKNDFFKSAMPBGYIDERTIFFQDDGKYTRREVAFEGDTLV 120
 QY 121 NRLELKGIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLELKGIDFKEDGNILGHKLEYNNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YGOONPILDGPVLPDNNHLYSTQSALSCKDPNEKRDHMYLLEFVTAAGITHGMDLYK 238
 DB 181 HYGTNPVLGDGPVLPDNNHLYSTQSALSCKDPNEKRDHMYLLEFVTAAGITHGMDLYK 238

RESULT 14

095UA7 PRELIMINARY; PRT; 225 AA.

AC 095UA7; PRELIMINARY; PRT; 225 AA.
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Cyan fluorescent protein (Fragment).
 OS Montastraea cavernosa (great star coral).
 OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Falkowski P.G., Sun Y.;
 RT "Montastraea cavernosa fluorescent protein."; Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AY056460; AAL17905.1;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP. 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 DR NON_TER 225
 FT SEQUENCE 225 AA; 25775 MW; 52DE2F716D083524 CRC64;
 SQ

Query Match 20.1%; Score 255.5; DB 5; Length 225;

Best Local Similarity 25.8%; Pred. No. 3.2e-13;

Matches 65; Conservative 49; Mismatches 69; Indels 69; Gaps 9;

QY 10 AVVPILVELDGVNGHKFSVSGEGSDVYTKLTKF-ICTTGKLPVMPPTLVTFPSYGV 68
 DB 6 SVAKIKLMQGIYNGHFMITGEGKPEEGTHTILKKEGGLPFAVDILTTAFQYGN 65
 QY 69 QCFSRYPDHMKRHFESKAMPBGVYQORTIFFKDDGNYKTRAEVKEGGDTLVNRLELKG 128
 DB 66 RVETKTPKDIIP--DYFKQSPPEGSYSWERSMTFEDQGVCTVSDIKLEGDCEFFYEIRFGV 123
 QY 129 DFEKEDGNILGHK-LEYNNSHNYIMADKOKNGI-----KVNFKIRHN 170
 DB 124 NFPSGVPVQKTKLKEPSTENMY-----RDGVLLGDVSRITLLBEDKHHRCNFRSTYG 178
 QY 171 IEDGSVOLADYQOONPILDGPVLPDNNHLYSTQSALSCKDPNEKRDHMYLLEFVTAAGIT 230
 DB 179 AKKG-----VVLPEYHFV-----DHRIE-----LS 199
 QY 231 HGMD---ELYK 238
 DB 200 HDKDYNTVEYE 211

RESULT 15

0963F5 PRELIMINARY; PRT; 225 AA.

AC 0963F5;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Green fluorescent protein.
 OS Montastraea cavernosa (great star coral).

OC Eukaryota; Metazoa; Chordata; Anthozoa; Zoantharia; Scleractinia;
 OC Faviina; Faviidae; Montastraea.
 OX NCBI_TaxID=63558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lesser M.P., Barry T.M., Mazel C., Matz M.V., Lukyanov S.A.,
 RA Falkowski P., Gorbunov M., Kolber Z.;
 RT "Green fluorescent proteins in Caribbean Scleractinian corals."; Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AF384683; AAK62982.2;
 DR InterPro; IPR000786; Green_fl_protein.
 DR Pfam; PF01353; GFP. 1.
 DR ProDom; PD013756; Green_fl_protein; 1.
 DR SEQUENCE 225 AA; 25847 MW; 77DE7D7C616929AF CRC64;
 SQ

Query Match 19.3%; Score 245; DB 5; Length 225;

Best Local Similarity 28.2%; Pred. No. 2.3e-12;

Matches 58; Conservative 47; Mismatches 89; Indels 12; Gaps 6;

QY 11 VVPILVELDGVNGHKFSVSGEGSDVYTKLTKFICTTG-KLPVMPPTLVTFPSYGV 69
 DB 7 IMEIKLMQGVNGHFMITGEGKPEEGTHTILKKEGGLPFAVDILTTAFQYGNR 66
 QY 70 CFSRYPDHMKRHFESKAMPBGVYQORTIFFKDDGNYKTRAEVKEGGDTLVNRLELKG 129
 DB 67 VETKTPKDIIP--DYFKQSPPEGSYSWERSMTFEDQGVCTVSDIKLEGDCEFFYEIRFGV 124
 QY 130 FKEKEDGNILGHK-LEYNNSHNYIMADKOKNGI-KVNFKIRHNIEDGSVOLADYQOONP 187
 DB 125 FPRNGVPVQKTKLKEPSTENMY-----RDGVLLGDVSRITLLBEDKHHRCNFRSTYG 177
 QY 188 ILDGPVLPDNNHLYSTQSALSCKDPNE 213
 DB 178 KAKKRVQLPDYHRYVDRHRIELSHND 203

Search completed: July 11, 2003, 17:58:45
 Job time : 83 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:51:18 ; Search time 23 Seconds

(without alignments)
429.190 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270
Sequence: 1 MSKGELEFTRVAVPILVELDG.....VLEFVTAGITGMDELYK 238

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt.40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1242	97.8	238	1	GFP_AEQVI
2	100.5	7.9	879	1	STL_XYLA
3	99	7.8	336	1	YD48_MENJA
4	90	7.1	795	1	D152_HAEIN
5	90	7.1	797	1	D151_HAEIN
6	89.5	7.0	393	1	MT04_HELPE
7	89	7.0	793	1	D153_HAEIN
8	88.5	7.0	312	1	TRXB_CHLMU
9	87.5	6.9	861	1	SYL_HAEIN
10	87	6.9	2334	1	MAPA_BACSU
11	86	6.8	357	1	TRMA_CAMJE
12	85.5	6.7	752	1	NEC1_RAT
13	85.5	6.7	886	1	TRH3_MESAU
14	84.5	6.7	312	1	TRXB_CHLMU
15	84.5	6.7	538	1	GRBE_RAT
16	82.5	6.5	613	1	PEPF_MYCPU
17	82.5	6.5	1164	1	BAG_STRAG
18	82	6.5	1433	1	DPO3_BACHD
19	81	6.4	836	1	DPO1_HPBUD
20	81	6.4	953	1	YMY7_YEAST
21	80.5	6.3	355	1	PLK_CHICK
22	80.5	6.3	504	1	MPPA_SOLUT
23	80.5	6.3	2222	1	DPOE_YEAST
24	80.5	6.3	6359	1	BACC_BACLI
25	80	6.3	774	1	AMT2_SCHPO
26	79.5	6.3	533	1	NIFD_CLOPA
27	79.5	6.3	887	1	TRH3_RAT
28	79.5	6.3	1224	1	COPA_HUMAN
29	79	6.2	339	1	PMAR_CANGA
30	79	6.2	354	1	PLK_HORSE
31	79	6.2	3135	1	S230_PLARO
32	78	6.1	217	1	EXPI_ERWCA
33	78	6.1	520	1	CC3_YEAST

34	78	6.1	664	1	SYM_BACSU	P27465 bacillus su
35	77.5	6.1	269	1	PENK_MOUSE	P22005 mus musculu
36	77.5	6.1	461	1	PSBC_CYAPA	P48104 cyanophora
37	77.5	6.1	540	1	NSA_MGCN	P75591 mycoplasma
38	77.5	6.1	682	1	PRC_ECOLI	P23865 escherichia
39	77.5	6.1	693	1	TGM3_HUMAN	Q08188 homo sapien
40	77.5	6.1	860	1	SYL_ECOLI	P07813 escherichia
41	77.5	6.1	1228	1	NARG_BACSU	P42175 bacillus su
42	77.5	6.1	1666	1	CO3_CAVPO	P12387 cavia porce
43	77	6.1	438	1	SHPI_STAPH	O08002 staphylococ
44	77	6.1	753	1	NEC1_MOUSE	P21662 mus musculu
45	76.5	6.0	220	1	TTM2_CAVPO	Q9WUC6 cavia porce

ALIGNMENTS

RESULT 1	ID	GFP_AEQVI	STANDARD:	PRT:	238 AA.
AC	P42212	017104:			
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	16-OCT-2001	(Rel. 40, Last annotation update)			
DE	Green fluorescent protein.				
GN	GFP.				
OS	Aequorea victoria (Jellyfish).				
OC	Eukaryota; Metazoa; Cnidaria; Hydrozoa; Hydroids; Leptomedusae;				
OX	Aequoreidae; Aequorea.				
NCBI_Taxid=6100;					
NP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.				
NP	MEDLINE=92175527; PubMed=1347277;				
RA	Prasher D.C., Eckenrode V.K., Ward W.W., Prendergast F.G.,				
RA	Cornlier M.J.,				
RT	"Primary structure of the Aequorea victoria green-fluorescent				
RT	protein.";				
RL	Gene 111:229-233(1992).				
NP	SEQUENCE FROM N.A.				
NP	MEDLINE=94185810; PubMed=8137953;				
RA	Inouye S., Tsuji F.I.,				
RT	"Aequorea green fluorescent protein. Expression of the gene and				
RT	fluorescence characteristics of the recombinant protein.";				
RL	FEBS Lett. 341:277-280(1994).				
NP	Biochemistry 32:1212-1218(1993).				
NP	CHROMOPHORE.				
NP	MEDLINE=93192221; PubMed=8448132;				
RA	Cody C.W., Prasher D.C., Westler W.W., Prendergast F.G., Ward W.W.,				
RT	"Chemical structure of the hexapeptide chromophore of the Aequorea				
RT	green-fluorescent protein.";				
RL	Biochemistry 32:1212-1218(1993).				
NP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
NP	MEDLINE=9635565; PubMed=8703075;				
RA	Orme M., Cubitt A.B., Kallio K., Gross L.A., Tsien R.Y.,				
RA	Remington S.J.,				
RT	"Crystal structure of the Aequorea victoria green fluorescent				
RT	protein.";				
RL	Science 273:1392-1395(1996).				
NP	X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).				
NP	MEDLINE=98294543; PubMed=9631087;				
RA	Yang F., Moss L.G., Phillips G.N. Jr.,				
RT	"The molecular structure of green fluorescent protein.";				
RL	Nat. Biotechnol. 14:1246-1251(1996).				
NP	X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF MUTANT WITH YELLOW EMISSION.				
NP	MEDLINE=98455509; PubMed=9782051;				
RA	Wachter R.M., Elsigler M.A., Kallio K., Hanson G.T., Remington S.J.,				
RT	"Structural basis of spectral shifts in the yellow-emission variants				
RT	of green fluorescent protein.";				
RL	Structure 6:1267-1277(1998).				

[7] X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
 RX MEDLINE=99238303; Pubmed=10220315;
 RA Esliger M.A., Wachter R.M., Hanson G.T., Kallio K., Remington S.J.;
 RT "Structural and spectral response of green fluorescent protein
 variants to changes in pH."
 RL Biochemistry 38:5296-5301(1999).
 CC -1- FUNCTION: ENERGY-TRANSFER ACCEPTOR. ITS ROLE IS TO TRANSDUCE THE
 BLUE CHEMILUMINESCENCE OF THE PROTEIN AECORIN INTO GREEN
 FLUORESCENT LIGHT BY ENERGY TRANSFER. FLUORESCES IN VIVO UPON
 RECEIVING ENERGY FROM THE CA(2+)-ACTIVATED PHOTOPROTEIN AECORIN.
 CC ABSORBS LIGHT MAXIMALLY AT 395 NM AND EXHIBITS A SMALLER
 CC ABSORPTION PEAK AT 470 NM. THE FLUORESCENCE EMISSION SPECTRUM
 CC PEAKS AT 509 NM WITH A SHOULDER AT 540 NM.
 CC -1- SUBUNIT: MONOMER.
 CC -1- TISSUE SPECIFICITY: PHOTOCYTES.
 CC -1- PPM: CONTAINS A COVALENTLY ATTACHED CHROMOPHORE, WHICH IS COMPOSED
 CC OF MODIFIED AMINO ACID RESIDUES. THE CHROMOPHORE IS FORMED UPON
 CC CYCLIZATION OF THE RESIDUES SER-DEHYDROTYR-GLY.
 CC -1- BIOTECNOLOGY: Has become a useful and ubiquitous tool for making
 CC chimeric proteins of GFP linked to other proteins where it
 CC functions as a fluorescent protein tag. GFP tolerates N- and C-
 CC terminal fusion to a broad variety of proteins. It has been
 CC expressed in bacteria, yeast, slime mold, plants, drosophila,
 CC zebrafish, and in mammalian cells. As a noninvasive fluorescent
 CC marker in living cells, it allows for a wide range of applications
 CC where it may function as a cell lineage tracer, reporter of gene
 CC expression, or as a measure of protein-protein interactions.
 CC -1- DATABASE: NAME=protein Spotlight;
 CC NOTE=Issue 11 of June 2001;
 CC WWW=ftp://www.expasy.org/spotlight/articles/spt1011.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M62654; AAA27722.1; -;
 DR EMBL: M62653; AAA27721.1; -;
 DR EMBL: L29345; AAA58246.1; -;
 DR PDB: 1GFL; 11-JAN-97.
 DR PDB: 1EMA; 08-NOV-96.
 DR PDB: 1EMA; 16-JUN-97.
 DR PDB: 1EMC; 20-AUG-97.
 DR PDB: 2EMD; 20-AUG-97.
 DR PDB: 1EMF; 20-AUG-97.
 DR PDB: 1EMF; 20-AUG-97.
 DR PDB: 1EMF; 12-MAY-99.
 DR PDB: 1EMK; 20-AUG-97.
 DR PDB: 1EML; 20-AUG-97.
 DR PDB: 1EMM; 20-AUG-97.
 DR PDB: 2EMN; 20-AUG-97.
 DR PDB: 2EMO; 20-AUG-97.
 DR PDB: 1BFP; 07-JUL-97.
 DR PDB: 1YFP; 28-OCT-98.
 DR InterPro: IPR000786; Green_fl_protein.
 DR Pfam: PF01353; GFP; 1.
 DR PRINTS: PRO1229; GLUORESCENT.
 DR ProDom: PD013756; Green_fl_protein; 1.
 KW Luminescence; 3D-structure.
 FT SITE 65 67 MODIFIED TO FORM THE CHROMOPHORE.
 FT VARIANT 100 100 F -> Y.
 FT VARIANT 108 108 T -> S.
 FT VARIANT 141 141 L -> M.
 FT VARIANT 219 219 V -> I.
 FT CONFLICT 25 25 H -> Q (IN REF. 2).
 FT CONFLICT 157 157 Q -> P (IN REF. 2).
 FT CONFLICT 172 172 E -> K (IN REF. 2).
 SO SEQUENCE 238 AA; 26886 MW; EA5A6F21FBB6E05 CRC64;

Query Match 97.8%; Score 1242; DB 1; Length 238;
 Best Local Similarity 97.5%; Pred. No. 4,4e-98;
 Matches 232; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MSKGEELFVAVPLVLDGVDGNGKFSVSGEGEDVTYKLIKFTCTGKLPVMPPTL 60
 DB 1 MSKGEELFVAVPLVLDGVDGNGKFSVSGEGEDVTYKLIKFTCTGKLPVMPPTL 60
 QY 61 VTFEFSVVOCESRPDMKRRHDFKSMAPGQVOORTFFKDDGNYTRAEVFEQSTLV 120
 DB 61 VTFEFSVVOCESRPDMKRRHDFKSMAPGQVOORTFFKDDGNYTRAEVFEQSTLV 120
 QY 121 NRTELKIDKREDGNILGHLREYNVNSHNYIMADKOKNCKIKVFKIRNIEDGSVOLAD 180
 DB 121 NRTELKIDKREDGNILGHLREYNVNSHNYIMADKOKNCKIKVFKIRNIEDGSVOLAD 180
 QY 181 YVOQNPFLDGPVLLPNNHTLSQALSADPNKRDHMLLEVTAAAGITHGDELYK 238
 DB 181 YVOQNPFLDGPVLLPNNHTLSQALSADPNKRDHMLLEVTAAAGITHGDELYK 238
 DB 181 HYQDNPFGDGPVLLPNNHTLSQALSADPNKRDHMLLEVTAAAGITHGDELYK 238
 RESULT 2
 SYL_XYLFA STANDARD; PRT; 879 AA.
 AC 09PBG8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
 GN LEU5 OR XF2176.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 CC NCBI_TaxID=2371.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9a5c;
 RX MEDLINE=20365717; Pubmed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Britones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carriro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferrao J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garner M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohnselt J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
 RA Menck C.F.M., Martins E.A.L., Martins E.M.F., Matsushima A.Y.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nanni A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Pelxoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pessegueiro J.B.,
 RA Ouagigo R.B., Roberto P.A.G., Rodrigues V. de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Savasati H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Silveira W.J., de Souza A.A.,
 RA de Souza A.P., Terezi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zait M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) -> AMP +
 CC diphosphate + L-leucyl-tRNA(Leu).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL--TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

RA Kyriades N.C., Moese C.R.;
RT "tetraarico-peptide-repeat proteins in the archaeon Methanococcus
RL Trends Biochem. Sci. 23:245-247(1998).
CC jannaschli."
CC
CC -1- CAUTION: According to Ref.2, this sequences contains TPR repeats.
CC These are not detected using our methodology.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/Announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; 067574; AAB99360.1; -.
CC DR TIGR; MJ1348; -.
CC KW Hypothetical protein; Complete proteome.
CC SQ SEQUENCE 336 AA; 40443 MW; 41BF2F28CDB036FE CRC64;
CC -----
Query Match 7.8%; Score 99; DB 1; Length 336;
Best Local Similarity 24.3%; Pred. No. 0.29; Indels 104; Gaps 19;
Matches 67; Conservative 40; Mismatches 65;

OY 19 DGDVNGHGFVSQGE-----GEGDVTYGKLTLEKTLCTTGKTLVPM-----57
DB 91 DGDV-----YNSGSLSTIASIFAKIKGLDI-----TNPLNSGGE-----WIYNDVINDAN 137
OY 58 -----PLVYTSST--GVGCFSPKPDHMKR-----HDFKSAPEGVYQORTIF 99
DB 138 SEDILKSVLTFEDFSEYERKEKLLNREP--HLRKLFEEDNIYTNFNSDFDEPM-----MF 187
OY 100 FKDDGNKVTAEVKEE-----GDTLVNRLTELKGIDPKEDGNIL--GKLENYNYS 147
DB 188 FIGAGNKKRFLLEVYEEFKNKIKKSOISNEYINEL--IKRPD--KMSDILAHLKLENYK 245
OY 148 HNYTIMADK-----OKNGIKVNFKIRHNIEDGV-----QLADYQOANTPIL 189
DB 246 CLYVYLMFEKFEYDFDEFNEIKKNKINILVDAIYMLKNGVAKKEWMLNRLNIYEKE--II 302
OY 190 DGPVLLPDNHYLSTQSAISKD--PREKPDHMYLLEFV 224
DB 303 KRP--LPNTY---KDAHNDLMLNELLIDYVYLKEFI 332

RESULT 4
D152_HAEMIN STANDARD: PRT; 795 AA.
AC P44935;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
DE (D-15-Rg) (Outer membrane protein D15).
GN H10917.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907.
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInerney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Gload A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Furumann J.L., Georgagagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae

```

RT Rd.":
 RL Science 269:496-512(1995).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32773; AAC2575.1; -
 DR TIGR: H10917; -
 DR InterPro: IPR000184; Bac.surfAg_D15.
 DR Pfam: PF01103; Bac.surfAg_1.
 KW Antigen; Outer membrane; Signal; Complete proteome.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 795 PROTECTIVE SURFACE ANTIGEN D15.
 SQ SEQUENCE 795 AA; 87478 MW; B85691FC22E6ED4 CRC64;
 Query Match 7.1%; Score 90; DB 1; Length 795;
 Best Local Similarity 20.2%; Pred. No. 4.8;
 Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;
 QY 22 VNGHK-----FSVSGEGEDVTYGLTLKFLCTTGKLPVWPPLVTTFSYGVOCFSRYP 75
 Db 402 INGSNDEVDVYKKEKNTGSINCG-----IGYFSGISYQ 438
 QY 76 DHKRRHDF-----FKSAMPEGYVOQ-----RTIEFKDGNNTYTRA 110
 Db 439 ASVKQDNFLGTGAASVAGTKNDYGVNLTGTEPYTKDGVSLGAVFPEFENDNSKSDT 498
 QY 111 EVKFEGLTLNRIELKIDFKEDGNI---LGH-----KLEYNSHWYIYADAKOK- 158
 Db 499 SSNWKRTTYSNVTL-GFPVNNNSYVVGIGHYNNKISNPALEN---RLUYIOSMKFKG 554
 QY 159 NGIKVN-FKIRHNIEDSGVOLADYYQONPTI-LDGPVLLP--DNHYLSTOSALSKDENEK 214
 Db 555 NGIKTNDPFSFGMNYNSLRNGYFPPTKGVKASLGGRVTIPGSDKKYKLSADVOGFYPLD 614
 QY 215 RDHNVLLFVTAGITG 232
 Db 615 RDHLWVVSARASAGYANG 632
 RESULT 5
 D15_HAEIN STANDARD; PRT; 797 AA.
 AC P46024;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D15-A9) (Outer membrane protein D15).
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
 OC Haemophilus.
 OC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Serotype B;
 RX MEDLINE=95255676; PubMed=7737523;
 RA Flack F.S., Loomore S., Chong P., Thomas W.R.;
 RT "The sequencing of the 80-kDa D15 protective surface antigen of
 RT Haemophilus influenzae";
 RL Gene 136:97-99(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Mina, and Eagan / Serotype B;
 RX MEDLINE=97427952; PubMed=9284140;
 RA Loomore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,

RA Klein M.H.;
 RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
 RT species and may represent a universal protective antigen against
 RT invasive disease";
 RL Infect. Immun. 65:3701-3707(1997).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U13961; AAA85645.1; -
 DR EMBL: U08032; AAB61974.1; -
 DR EMBL: U08033; AAB61976.1; -
 DR InterPro: IPR000184; Bac.surfAg_D15.
 DR Pfam: PF01103; Bac.surfAg_1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 797 PROTECTIVE SURFACE ANTIGEN D15.
 SQ SEQUENCE 797 AA; 87675 MW; 2F93DE538696AFLB CRC64;
 Query Match 7.1%; Score 90; DB 1; Length 797;
 Best Local Similarity 20.2%; Pred. No. 4.8;
 Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;
 QY 22 VNGHK-----FSVSGEGEDVTYGLTLKFLCTTGKLPVWPPLVTTFSYGVOCFSRYP 75
 Db 402 INGSNDEVDVYKKEKNTGSINCG-----IGYFSGISYQ 438
 QY 76 DHKRRHDF-----FKSAMPEGYVOQ-----RTIEFKDGNNTYTRA 110
 Db 439 ASVKQDNFLGTGAASVAGTKNDYGVNLTGTEPYTKDGVSLGAVFPEFENDNSKSDT 498
 QY 111 EVKFEGLTLNRIELKIDFKEDGNI---LGH-----KLEYNSHWYIYADAKOK- 158
 Db 499 SSNWKRTTYSNVTL-GFPVNNNSYVVGIGHYNNKISNPALEN---RLUYIOSMKFKG 554
 QY 159 NGIKVN-FKIRHNIEDSGVOLADYYQONPTI-LDGPVLLP--DNHYLSTOSALSKDENEK 214
 Db 555 NGIKTNDPFSFGMNYNSLRNGYFPPTKGVKASLGGRVTIPGSDKKYKLSADVOGFYPLD 614
 QY 215 RDHNVLLFVTAGITG 232
 Db 615 RDHLWVVSARASAGYANG 632
 RESULT 6
 MT04_HELPY STANDARD; PRT; 393 AA.
 ID MT04_HELPY
 AC O25443;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical methyltransferase HP0747 (Ec 2.1.1.-).
 GN HP0747
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OC NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=97394467; PubMed=9252185;
 RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khakhria H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,

RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
 RA Cotton M.D., Meldman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.:
 "The complete genome sequence of the gastric pathogen *Helicobacter*
 RT *pylori*.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: PROBABLE METHYLTRANSFERASE.
 CC -1- SIMILARITY: BELONGS TO THE UPF0155 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AE000587; AAD07796.1; -
 DR TIGR; HP0747; -
 DR InterPro: IPR004395; Cons_hypoth91.
 DR InterPro: IPR003358; Methyltransf_4.
 DR InterPro: IPR000051; SAM_Bind.
 DR Pfam; PF02390; Methyltransf_4; 1.
 DR TIGRFAMs; TIGR00091; Cons_hypoth91; 1.
 KW Hypothetical protein; Transferase; Methyltransferase;
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 45774 MW; C376EF02D33F71D1 CRC64;

Query Match 7.0%; Score 89.5; DB 1; Length 393;
 Best Local Similarity 20.8%; Pred. No. 2.3;
 Matches 51; Conservative 42; Mismatches 77; Indels 75; Gaps 15;

QY 34 EGDVYTGKLT-----FICTGKLPVMPVLTFTSYGQCSRPDMMKR-- 81
 DB 174 QGD---GRVLESMPNRCERLIV---HFPVPW-----MEKKRR 207
 QY 82 ----DFKFSAM-----PEGYVOQR--IFFKDD-----GNKTRAEVKFEQD-TIVNRIE 124
 DB 208 VLSKFLNALRVKPRGFLRLRDTDSIFEDSLKLNKQCELEIKNMQIPVSKYE 267
 QY 125 LKGIQKED-GNIGHKLEIV---YNSHNYIMADK--QRNGIKVNEKIRHNIEDG-SV 176
 DB 268 ARWKKLKLDIYDLRIYSLENNETPFEDNAFSEFYTITISKSVGTIIKTKKIIQEGYFV 327
 QY 177 QLADYVOQNTPI-----DGR-----VLLPDNHYLSTQSALSDPBEKDDHWLLEFV 224
 DB 328 HVCNITYENKGFIVELSKGDFWVPRFLVLTENOIFYLNKSPKLTNNHKAHLLONTL 387
 QY 225 TAAGI 229
 DB 388 SQKGI 392

RESULT 7
 ID D153_HAEIN STANDARD: PRT; 793 AA.
 AC 032629;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protective surface antigen D15 precursor (80 kDa D15 antigen)
 DE (D-15-Ag) (Outer membrane protein D15).
 OS Haemophilus influenzae.
 CC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus.
 OC NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PAK 12085;
 RX MEDLINE=97427952; PubMed=9284140;
 RA Loosmore S.M., Yang Y.P., Coleman D.C., Shortreed J.M., England D.M.,
 RA Klein M.H.;

RT "Outer membrane protein D15 is conserved among Haemophilus influenzae
 RT species and may represent a universal protective antigen against
 RT invasive disease.";
 RL Infect. Immun. 65:3701-3707(1997).
 CC -1- SUBCELLULAR LOCATION: Outer membrane.
 CC -1- SIMILARITY: BELONGS TO THE SURFACE ANTIGEN D15 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U60834; AA61977.1; -
 DR InterPro: IPR000184; Bac_surfAg_D15.
 DR Pfam; PF01103; Bac_surface_Ag; 1.
 KW Antigen; Outer membrane; Signal.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 793 PROTECTIVE SURFACE ANTIGEN D15.
 SQ SEQUENCE 793 AA; 87511 MW; 51BFB203680A1A4 CRC64;

Query Match 7.0%; Score 89; DB 1; Length 793;
 Best Local Similarity 20.5%; Pred. No. 5.8;
 Matches 53; Conservative 34; Mismatches 97; Indels 74; Gaps 12;

QY 22 VNGHK-----FVSGBEGDYTGKLTFRCTTGKLPVMPVLTFTSYGQCSFRRP 75
 DB 402 INGSNDEVDVYKVKERNWTSINFG-----IGGTSGISYQ 438
 QY 76 DHMKRHF-----FKSAMPEGYVOO-----RTIFFKDDGKYTRA 110
 DB 439 TSIKODNLTGCAAVSINAGTNDGTSVNLGTEFTFKDGVSLGAGNIFENYDNSKDT 498
 QY 111 EVKEGDTLVRIELKGIQKEDGNI---LGH-----KLEYNNSHNYIMADKOK 158
 DB 499 SSNKRTTYGSNVTL-GPVENNSYVYGLGHTYKKNISFALDEVN---RNLIQSMKRFK 554
 QY 159 NGIKVY-KIRHNIEDGSVOADYQOQNTPI-LDGVPVLLP--DNHYLSTQSALSDPBEK 214
 DB 555 NGIKINDDFSGFWMYNSLNGRYPTKGVKASLGGRVYIPSDMKYKRLSDVGGFYPLD 614
 QY 215 RDHWVLEFVTAAGITGH 232
 DB 615 RDHRVYVSAKASAGYANG 632

RESULT 8
 ID TRXB_CHLMU STANDARD: PRT; 312 AA.
 AC 09PKT7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioresoxin reductase (EC 1.6.4.5) (TRXR).
 GN TRXB OR TC0375.
 OS Chlamydia muridarum.
 CC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OC NCBI_TaxID=83560;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MOPn / N19g;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.R., Peterson J., Utterback T., Berry K., Baas S.,
 RA Linher K., Weidman J., Khouiri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 "Genome sequences of Chlamydia trachomatis MOPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 CC -1- CATALYTIC ACTIVITY: NADPH + oxidized thioresoxin = NADP(+) +

```

CC reduced thioredoxin.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
CC OXIDOREDUCTASES CLASS-II.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002304; AAF39233.1; -.
DR HSSP: Q39243; IYDC.
DR TIGR: TC0375; -.
DR InterPro: IPR000759; Adrndx_reductase.
DR InterPro: IPR001327; FAD_pyr_redox.
DR InterPro: IPR001100; Pyr_redox.
DR Pfam: PF00070; Pyr_redox_1.
DR PRINTS: PR00419; ADXRORASE.
DR PRINTS: PR00368; FADPNR.
DR PRINTS: PR00411; PNDRTASEI.
DR PRINTS: PR00469; PNDRTASEI.
DR PRODOM: PD000139; FAD_pyr_redox; 1.
DR TIGRFAMS: TIGR01292; TRX_reduct; 1.
DR PROSITE: PS00573; PYRIDINE_REDOX_2; 1.
DR Redox-active center: Oxidoreductase; NADP; Flavoprotein; FAD;
KW Complete proteome.
FT NP_BIND 4 19 FAD (ADP PART) (PROBABLE).
FT DISULFID 138 141 REDOX-ACTIVE (BY SIMILARITY).
FT NP_BIND 152 166 NAD(P) (BY SIMILARITY).
FT NP_BIND 273 283 FAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 312 AA; 33556 MW; 53824B207C2A155D CRC64;

Query Match
Best Local Similarity 7.0%; Score 88.5; DB 1; Length 312;
Matches 48; Conservative 29; Mismatches 81; Indels 69; Gaps 9;

QY 1 MSGGEELFVAVPIVLELDGVNGHKSVSSEGGSDVYTKLTKFCTTGKLPVWPTL 60
DB 97 LKSGEDFTCDACIAT--GASAKRLSIPGAGDNEFMOKGVACAVC-DGASPI----- 147
QY 61 VTTFSSVGCFSRYPDMKRRHDF-----KSAMEGVVQORTIFPKDGNVKTAEYKFE 115
DB 148 -----FRDRDLFVIGGGSALAEAFILR-----YGRVYVYHR 181
QY 116 GDPL-----VNRLEKIDGFKEDGNILGHKLEYNNSHNYITMADRKNGIKVYFKIRH 169
DB 182 RDLFRASKAMVKNQAQ-----ANEKIVFLMNSVYVKILGDSLVASI-----DIFN 225
QY 170 NIDEGSVQL-----ADYVQONTPLDGPVLLPDMNYITLSTQSALSK 209
DB 226 NVEKTYTMEAGVFAIGHQPTAFLGQSLDENGYYITTEKSSR 272

RESULT 9
SYL_HAEIN STANDARD; PRT; 861 AA.
AC P43827;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine--tRNA ligase) (LeuRS).
GN LEUCS OR HI0921.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kesteven K., Sutton G., Fitzhugh W., Fields C.A., Goodyne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Leu) -> AMP +
CC diphosphate + L-leucyl-tRNA(Leu).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33774; AAC22581.1; -.
DR TIGR: HI0921; -.
DR InterPro: IPR002302; Leu-tRNA-synt.1a.
DR InterPro: IPR002300; tRNA-synt.1a.
DR InterPro: IPR001412; tRNA-synt.1.
DR Pfam: PF00133; tRNA-synt.1; 1.
DR PRINTS: PR00985; TRNASYNTLEU.
DR TIGRFAMS: TIGR00396; LeuS_bact; 1.
DR PROSITE: PS00178; AA_tRNA_LIGASE_1; 1.
DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 42 52 "HIGH" REGION.
FT SITE 619 623 "KMSKS" REGION.
FT BINDING 622 622 ATP (BY SIMILARITY).
SQ SEQUENCE 861 AA; 97750 MW; EB93304F6B4C8FB7 CRC64;

Query Match
Best Local Similarity 6.9%; Score 87.5; DB 1; Length 861;
Matches 45; Conservative 25; Mismatches 68; Indels 47; Gaps 9;

QY 49 TTGKLPVWPTLVTTFSYVGCFSRYPDMKRRHDFKSAPEGVVQORTIFPKDGNVKT 108
DB 314 TGDKLPI-WVANFVLMHYGTGAWAAPAHQORD--FEFAQKYSTPIKQVIAPIADERIDL 370
QY 109 RAEYKPEGDLVNRLEKIDGFKEDGNILGHKLEYNNSHNYITMADRKNGIKVYFKIRH 164
DB 371 TKQAEVHGKLVNSDEDFGNF--DGAENG-----LADKLEKLVGKRQVN 414
QY 165 FKIRH-----NIDEGSVQLADYVQONTPLDGPVLLPDMNYITLSTQSAL 207
DB 415 YRLRDMGVSRQRYWGAPIPMLITLENGDVYRA-----PREDLPILPELVNMDGVKSP 467
QY 208 SKDPN 212
DB 468 NADPN 472

RESULT 10
WAPA_BACSU STANDARD; PRT; 2334 AA.
ID WAPA_BACSU
AC Q07833;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Wall-associated protein precursor.
 GN WAPA OR N17G.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NC NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=93302506; PubMed=8316082;
 RA Foster S.J.;
 RT "Molecular analysis of three major wall-associated proteins of
 Bacillus subtilis 168: evidence for processing of the product of a
 gene encoding a 258 kDa precursor two-domain ligand-binding
 protein."
 RL Mol. Microbiol. 8:299-310(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=95219088; PubMed=7704263;
 RA Yoshida K.-I., Sano H., Seki S., Oda M., Fujimura M., Fujita Y.;
 RT "Cloning and sequencing of a 29 kb region of the Bacillus subtilis
 genome containing the hut and wapa loci."
 RL Microbiology 141:337-343(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168 / BGSC1A1;
 RX MEDLINE=97124196; PubMed=8969509;
 RA Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,
 Miwa Y., Fujita Y.;
 RT "Sequencing of a 65 kb region of the Bacillus subtilis genome
 containing the lic and cel loci, and creation of a 177 kb contig
 covering the gut-saxy region."
 RL Microbiology 142:3113-3123(1996).
 CC -1- FUNCTION: STILL UNKNOWN. NOT INVOLVED IN CELL MEMBRANE METABOLISM,
 MOTILITY, SECRETION OR DIFFERENTIATION.
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-ASSOCIATED. MAY BE RELEASED
 INTO THE MEDIUM.
 CC -1- DOMAIN: HAS TWO LIGAND-BINDING DOMAINS: THE N-TERMINUS, HAS THREE
 101 AA REPEATS WHICH ARE RESPONSIBLE FOR CELL WALL BINDING; THE
 C-TERMINUS CONSISTS OF TWO BLOCKS OF RESIDUES WITH A CONSERVED
 MOTIF REPEATED 31 TIMES.
 CC -1- SIMILARITY: THE REPEATED MOTIF IN THE C-TERMINUS HAS SOME
 SIMILARITY TO THE REPEAT IN E. COLI RBS GROUP OF PROTEINS (RBSA-D).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L05634; AAA22883.1; -;
 DR EMBL; D31856; BAA0656.1; -;
 DR EMBL; D29985; BAA06260.1; -;
 DR EMBL; D83026; BAA11683.1; -;
 DR EMBL; Z99124; CAB15959.1; -;
 DR PIR; S32920; S32920.
 DR Subtilist; BG10797; wapa.
 DR InterPro; IPR003305; CBM_4_9; 1.
 KW Pfam: PF02018; CBM_4_9; 1.
 KW Cell wall; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 28 OR 32 (POTENTIAL).
 FT CHAIN 29 2334 WALL-ASSOCIATED PROTEIN.
 FT DOMAIN 504 869 3 X 101 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 504 605 1-1.
 FT REPEAT 636 736 1-2.
 FT REPEAT 769 869 1-3.
 FT DOMAIN 1021 2139 31 X 21 AA APPROXIMATE TANDEM REPEATS OF
 FT REPEAT 1021 1040 X(4)-G-X(4)-[YF]-X-D-X(2)-G-X(4).
 FT REPEAT 1042 1061 2-1.
 FT REPEAT 2-2.

FT REPEAT 1063 1082 2-3.
 FT REPEAT 1083 1102 2-4.
 FT REPEAT 1109 1128 2-5.
 FT REPEAT 1129 1148 2-6.
 FT REPEAT 1150 1169 2-7.
 FT REPEAT 1174 1193 2-8.
 FT REPEAT 1199 1218 2-9.
 FT REPEAT 1219 1238 2-10.
 FT REPEAT 1246 1265 2-11.
 FT REPEAT 1267 1286 2-12.
 FT REPEAT 1290 1309 2-13.
 FT REPEAT 1312 1331 2-14.
 FT REPEAT 1332 1351 2-15.
 FT REPEAT 1373 1392 2-16.
 FT REPEAT 1414 1433 2-17.
 FT REPEAT 1440 1459 2-18.
 FT REPEAT 1480 1499 2-19.
 FT REPEAT 1520 1539 2-20.
 FT REPEAT 1561 1580 2-21.
 FT REPEAT 1602 1621 2-22.
 FT REPEAT 1643 1662 2-23.
 FT REPEAT 1684 1703 2-24 (APPROXIMATE).
 FT REPEAT 1725 1744 2-25.
 FT REPEAT 1787 1806 2-26.
 FT REPEAT 1828 1847 2-27.
 FT REPEAT 1869 1888 2-28.
 FT REPEAT 1910 1929 2-29.
 FT REPEAT 1951 1970 2-30.
 FT REPEAT 1992 2011 2-31.
 FT REPEAT 2033 2052 2-32.
 FT REPEAT 2074 2093 2-33.
 FT REPEAT 2115 2134 2-34.
 FT REPEAT 2156 2175 2-35.
 FT REPEAT 2197 2216 2-36.
 FT REPEAT 2238 2257 2-37.
 FT REPEAT 2279 2298 2-38.
 FT REPEAT 2320 2339 2-39.
 FT REPEAT 2361 2380 2-40.
 FT REPEAT 2402 2421 2-41.
 FT REPEAT 2443 2462 2-42.
 FT REPEAT 2484 2503 2-43.
 FT REPEAT 2525 2544 2-44.
 FT REPEAT 2566 2585 2-45.
 FT REPEAT 2607 2626 2-46.
 FT REPEAT 2648 2667 2-47.
 FT REPEAT 2689 2708 2-48.
 FT REPEAT 2730 2749 2-49.
 FT REPEAT 2771 2790 2-50.
 FT REPEAT 2812 2831 2-51.
 FT REPEAT 2853 2872 2-52.
 FT REPEAT 2894 2913 2-53.
 FT REPEAT 2935 2954 2-54.
 FT REPEAT 2976 2995 2-55.
 FT REPEAT 3017 3036 2-56.
 FT REPEAT 3058 3077 2-57.
 FT REPEAT 3099 3118 2-58.
 FT REPEAT 3140 3159 2-59.
 FT REPEAT 3181 3200 2-60.
 FT REPEAT 3222 3241 2-61.
 FT REPEAT 3263 3282 2-62.
 FT REPEAT 3304 3323 2-63.
 FT REPEAT 3345 3364 2-64.
 FT REPEAT 3386 3405 2-65.
 FT REPEAT 3427 3446 2-66.
 FT REPEAT 3488 3507 2-67.
 FT REPEAT 3529 3548 2-68.
 FT REPEAT 3580 3600 2-69.
 FT REPEAT 3621 3640 2-70.
 FT REPEAT 3662 3681 2-71.
 FT REPEAT 3703 3722 2-72.
 FT REPEAT 3744 3763 2-73.
 FT REPEAT 3785 3804 2-74.
 FT REPEAT 3826 3845 2-75.
 FT REPEAT 3867 3886 2-76.
 FT REPEAT 3908 3927 2-77.
 FT REPEAT 3949 3968 2-78.
 FT REPEAT 3990 4009 2-79.
 FT REPEAT 4031 4050 2-80.
 FT REPEAT 4072 4091 2-81.
 FT REPEAT 4113 4132 2-82.
 FT REPEAT 4154 4173 2-83.
 FT REPEAT 4195 4214 2-84.
 FT REPEAT 4236 4255 2-85.
 FT REPEAT 4277 4296 2-86.
 FT REPEAT 4318 4337 2-87.
 FT REPEAT 4359 4378 2-88.
 FT REPEAT 4400 4419 2-89.
 FT REPEAT 4441 4460 2-90.
 FT REPEAT 4482 4501 2-91.
 FT REPEAT 4523 4542 2-92.
 FT REPEAT 4564 4583 2-93.
 FT REPEAT 4605 4624 2-94.
 FT REPEAT 4646 4665 2-95.
 FT REPEAT 4687 4706 2-96.
 FT REPEAT 4728 4747 2-97.
 FT REPEAT 4769 4788 2-98.
 FT REPEAT 4810 4829 2-99.
 FT REPEAT 4851 4870 3-00.
 FT REPEAT 4892 4911 3-01.
 FT REPEAT 4933 4952 3-02.
 FT REPEAT 4974 4993 3-03.
 FT REPEAT 5015 5034 3-04.
 FT REPEAT 5056 5075 3-05.
 FT REPEAT 5097 5116 3-06.
 FT REPEAT 5138 5157 3-07.
 FT REPEAT 5179 5198 3-08.
 FT REPEAT 5220 5239 3-09.
 FT REPEAT 5261 5280 3-10.
 FT REPEAT 5302 5321 3-11.
 FT REPEAT 5343 5362 3-12.
 FT REPEAT 5384 5403 3-13.
 FT REPEAT 5425 5444 3-14.
 FT REPEAT 5486 5505 3-15.
 FT REPEAT 5527 5546 3-16.
 FT REPEAT 5568 5587 3-17.
 FT REPEAT 5609 5628 3-18.
 FT REPEAT 5650 5669 3-19.
 FT REPEAT 5691 5710 3-20.
 FT REPEAT 5732 5751 3-21.
 FT REPEAT 5773 5792 3-22.
 FT REPEAT 5814 5833 3-23.
 FT REPEAT 5855 5874 3-24.
 FT REPEAT 5896 5915 3-25.
 FT REPEAT 5937 5956 3-26.
 FT REPEAT 5978 5997 3-27.
 FT REPEAT 6019 6038 3-28.
 FT REPEAT 6060 6079 3-29.
 FT REPEAT 6101 6120 3-30.
 FT REPEAT 6142 6161 3-31.
 FT REPEAT 6183 6202 3-32.
 FT REPEAT 6224 6243 3-33.
 FT REPEAT 6265 6284 3-34.
 FT REPEAT 6306 6325 3-35.
 FT REPEAT 6347 6366 3-36.
 FT REPEAT 6388 6407 3-37.
 FT REPEAT 6429 6448 3-38.
 FT REPEAT 6470 6489 3-39.
 FT REPEAT 6511 6530 3-40.
 FT REPEAT 6552 6571 3-41.
 FT REPEAT 6593 6612 3-42.
 FT REPEAT 6634 6653 3-43.
 FT REPEAT 6675 6694 3-44.
 FT REPEAT 6716 6735 3-45.
 FT REPEAT 6757 6776 3-46.
 FT REPEAT 6798 6817 3-47.
 FT REPEAT 6839 6858 3-48.
 FT REPEAT 6880 6899 3-49.
 FT REPEAT 6921 6940 3-50.
 FT REPEAT 6962 6981 3-51.
 FT REPEAT 7003 7022 3-52.
 FT REPEAT 7044 7063 3-53.
 FT REPEAT 7085 7104 3-54.
 FT REPEAT 7126 7145 3-55.
 FT REPEAT 7167 7186 3-56.
 FT REPEAT 7208 7227 3-57.
 FT REPEAT 7249 7268 3-58.
 FT REPEAT 7290 7309 3-59.
 FT REPEAT 7331 7350 3-60.
 FT REPEAT 7372 7391 3-61.
 FT REPEAT 7413 7432 3-62.
 FT REPEAT 7454 7473 3-63.
 FT REPEAT 7495 7514 3-64.
 FT REPEAT 7536 7555 3-65.
 FT REPEAT 7577 7596 3-66.
 FT REPEAT 7618 7637 3-67.
 FT REPEAT 7659 7678 3-68.
 FT REPEAT 7700 7719 3-69.
 FT REPEAT 7741 7760 3-70.
 FT REPEAT 7782 7801 3-71.
 FT REPEAT 7823 7842 3-72.
 FT REPEAT 7864 7883 3-73.
 FT REPEAT 7905 7924 3-74.
 FT REPEAT 7946 7965 3-75.
 FT REPEAT 7987 8006 3-76.
 FT REPEAT 8028 8047 3-77.
 FT REPEAT 8069 8088 3-78.
 FT REPEAT 8110 8129 3-79.
 FT REPEAT 8151 8170 3-80.
 FT REPEAT 8192 8211 3-81.
 FT REPEAT 8233 8252 3-82.
 FT REPEAT 8274 8293 3-83.
 FT REPEAT 8315 8334 3-84.
 FT REPEAT 8356 8375 3-85.
 FT REPEAT 8397 8416 3-86.
 FT REPEAT 8438 8457 3-87.
 FT REPEAT 8479 8498 3-88.
 FT REPEAT 8520 8539 3-89.
 FT REPEAT 8561 8580 3-90.
 FT REPEAT 8602 8621 3-91.
 FT REPEAT 8643 8662 3-92.
 FT REPEAT 8684 8703 3-93.
 FT REPEAT 8725 8744 3-94.
 FT REPEAT 8766 8785 3-95.
 FT REPEAT 8807 8826 3-96.
 FT REPEAT 8848 8867 3-97.
 FT REPEAT 8889 8908 3-98.
 FT REPEAT 8930 8949 3-99.
 FT REPEAT 8971 8990 4-00.
 FT REPEAT 9012 9031 4-01.
 FT REPEAT 9053 9072 4-02.
 FT REPEAT 9094 9113 4-03.
 FT REPEAT 9135 9154 4-04.
 FT REPEAT 9176 9195 4-05.
 FT REPEAT 9217 9236 4-06.
 FT REPEAT 9258 9277 4-07.
 FT REPEAT 9299 9318 4-08.
 FT REPEAT 9340 9359 4-09.
 FT REPEAT 9381 9400 4-10.
 FT REPEAT 9422 9441 4-11.
 FT REPEAT 9463 9482 4-12.
 FT REPEAT 9504 9523 4-13.
 FT REPEAT 9545 9564 4-14.
 FT REPEAT 9586 9605 4-15.
 FT REPEAT 9627 9646 4-16.
 FT REPEAT 9668 9687 4-17.
 FT REPEAT 9709 9728 4-18.
 FT REPEAT 9750 9769 4-19.
 FT REPEAT 9791 9810 4-20.
 FT REPEAT 9832 9851 4-21.
 FT REPEAT 9873 9892 4-22.
 FT REPEAT 9914 9933 4-23.
 FT REPEAT 9955 9974 4-24.
 FT REPEAT 9996 10015 4-25.
 FT REPEAT 10037 10056 4-26.
 FT REPEAT 10078 10097 4-27.
 FT REPEAT 10119 10138 4-28.
 FT REPEAT 10160 10179 4-29.
 FT REPEAT 10201 10220 4-30.
 FT REPEAT 10242 10261 4-31.
 FT REPEAT 10283 10302 4-32.
 FT REPEAT 10324 10343 4-33.
 FT REPEAT 10365 10384 4-34.
 FT REPEAT 10406 10425 4-35.
 FT REPEAT 10447 10466 4-36.
 FT REPEAT 10488 10507 4-37.
 FT REPEAT 10529 10548 4-38.
 FT REPEAT 10570 10589 4-39.
 FT REPEAT 10611 10630 4-40.
 FT REPEAT 10652 10671 4-41.
 FT REPEAT 10693 10712 4-42.
 FT REPEAT 10734 10753 4-43.
 FT REPEAT 10775 10794 4-44.
 FT REPEAT 10816 10835 4-45.
 FT REPEAT 10857 10876 4-46.
 FT REPEAT 10898 10917 4-47.
 FT REPEAT 10939 10958 4-48.
 FT REPEAT 10980 11000 4-49.
 FT REPEAT 11021 11040 4-50.
 FT REPEAT 11062 11081 4-51.
 FT REPEAT 11103 11122 4-52.
 FT REPEAT 11144 11163 4-53.
 FT REPEAT 11185 11204 4-54.
 FT REPEAT 11226 11245 4-55.
 FT REPEAT 11267 11286 4-56.
 FT REPEAT 11308 11327 4-57.
 FT REPEAT 11349 11368 4-58.
 FT REPEAT 11390 11409 4-59.
 FT REPEAT 11431 11450 4-60.
 FT REPEAT 11472 11491 4-61.
 FT REPEAT 11513 11532 4-62.
 FT REPEAT 11554 11573 4-63.
 FT REPEAT 11595 11614 4-64.
 FT REPEAT 11636 11655 4-65.
 FT REPEAT 11677 11696 4-66.
 FT REPEAT 11718 11737 4-67.
 FT REPEAT 11759 11778 4-68.
 FT REPEAT 11800 11819 4-69.
 FT REPEAT 11841 11860 4-70.
 FT REPEAT 11882 11901 4-71.
 FT REPEAT 11923 11942 4-72.
 FT REPEAT 11964 11983 4-73.
 FT REPEAT 12005 12024 4-74.
 FT REPEAT 12046 12065 4-75.
 FT REPEAT 12087 12106 4-76.
 FT REPEAT 12128 12147 4-77.
 FT REPEAT 12169 12188 4-78.
 FT REPEAT 12210 12229 4-79.
 FT REPEAT 12251 12270 4-80.
 FT REPEAT 12292 12311 4-81.
 FT REPEAT 12333 12352 4-82.
 FT REPEAT 12374 12393 4-83.
 FT REPEAT 12415 12434 4-84.
 FT REPEAT 12456 12475 4-85.
 FT REPEAT 12497 12516 4-86.
 FT REPEAT 12538 12557 4-87.
 FT REPEAT 12579 12598 4-88.
 FT REPEAT 12620 12639 4-89.
 FT REPEAT 12661 12680 4-90.
 FT REPEAT 12702 12721 4-91.
 FT REPEAT 12743 12762 4-92.
 FT REPEAT 12784 12803 4-93.
 FT REPEAT 12825 12844 4-94.
 FT REPEAT 12866 12885 4-95.
 FT REPEAT 12907 12926 4-96.
 FT REPEAT 12948 12967 4-97.
 FT REPEAT 12989 13008 4-98.
 FT REPEAT 13030 13049 4-99.
 FT REPEAT 13071 13090 5-00.
 FT REPEAT 13112 13131 5-01.
 FT REPEAT 13153 13172 5-02.
 FT REPEAT 13194 13213 5-03.
 FT REPEAT 13235 13254 5-04.
 FT REPEAT 13276 13295 5-05.
 FT REPEAT 13317 13336 5-06.
 FT REPEAT 13358 13377 5-07.
 FT REPEAT 13399 13418 5-08.
 FT REPEAT 13440 13459 5-09.
 FT REPEAT 13481 13500 5-10.
 FT REPEAT 13522 13541 5-11.
 FT REPEAT 13563 13582 5-12.
 FT REPEAT 13604 13623 5-13.
 FT REPEAT 13645 13664 5-14.
 FT REPEAT 13686 13705 5-15.
 FT REPEAT 13727 13746 5-16.
 FT REPEAT 13768 13787 5-17.
 FT REPEAT 13809 13828 5-18.
 FT REPEAT 13850 13869 5-19.
 FT REPEAT 13891 13910 5-20.
 FT REPEAT 13932 13951 5-21.
 FT REPEAT 13973 13992 5-22.
 FT REPEAT 14014 14033 5-23.
 FT REPEAT 14055 14074 5-24.
 FT REPEAT 14096 14115 5-25.
 FT REPEAT 14137 14156 5-26.
 FT REPEAT 14178 14197 5-27.
 FT REPEAT 14219 14238 5-28.
 FT REPEAT 14260 14279 5-29.
 FT REPEAT 14301 14320 5-30.
 FT REPEAT 14342 14361 5-31.
 FT REPEAT 14383 14402 5-32.
 FT REPEAT 14424 14443 5-33.
 FT REPEAT 14465 14484 5-34.
 FT REPEAT 14506 14525 5-35.
 FT REPEAT 14547 14566 5-36.
 FT REPEAT 14588 14607 5-37.
 FT REPEAT 14629 14648 5-38.
 FT REPEAT 14670 14689 5-39.
 FT REPEAT 14711 14730 5-40.
 FT REPEAT 14752 14771 5-41.
 FT REPEAT 14793 14812 5-42.
 FT REPEAT 14834 14853 5-43.
 FT REPEAT 14875 14894 5-44.
 FT REPEAT 14916 14935 5-45.
 FT REPEAT 14957 14976 5-46.
 FT REPEAT 14998 15017 5-47.
 FT REPEAT 15039 15058 5-48.
 FT REPEAT 15080 15099 5-49.
 FT REPEAT 15121 15140 5-50.
 FT REPEAT 15162 15181 5-51.
 FT REPEAT 15203 15222 5-52.
 FT REPEAT 15244 15263 5-53.
 FT REPEAT 15285 15304 5-54.
 FT REPEAT 15326 15345 5-55.
 FT REPEAT 15367 15386 5-56.
 FT REPEAT 15408 15427 5-57.
 FT REPEAT 15449 15468 5-58.
 FT REPEAT 15490 15509 5-59.
 FT REPEAT 15531 15550 5-60.
 FT REPEAT 15572 15591 5-61.
 FT REPEAT 15613 15632 5-62.
 FT REPEAT 15654 15673 5-63.
 FT REPEAT 15695 15714 5-64.
 FT REPEAT 15736 15755 5-65.
 FT REPEAT 15777 15796 5-66.
 FT REPEAT 15818 15837 5-67.
 FT REPEAT 15859 15878 5-68.
 FT REPEAT 15900 15919 5-69.
 FT REPEAT 15941 15960 5-70.
 FT REPEAT 15982 16001 5-71.
 FT REPEAT 16023 16042 5-72.
 FT REPEAT 16064 16083 5-73.
 FT REPEAT 16105 16124 5-74.
 FT REPEAT 16146 16165 5-75.
 FT REPEAT 16187 16206 5-76.
 FT REPEAT 16228 16247 5-77.
 FT REPEAT 16269 16288 5-78.
 FT REPEAT 16310 16329 5-79.
 FT REPEAT 16351 16370 5-80.
 FT REPEAT 16392 16411 5-81.
 FT REPEAT 16433 16452 5-82.
 FT REPEAT 16474 16493 5-83.
 FT REPEAT 16515 16534 5-84.
 FT REPEAT 16556 16575 5-85.
 FT REPEAT 16597 16616 5-86.
 FT REPEAT 16638 16657 5-87.
 FT REPEAT 16679 16698 5-88.
 FT REPEAT 16720 16739 5-89.
 FT REPEAT 16761 16780 5-90.
 FT REPEAT 16802 16821 5-91.
 FT REPEAT 16843 16862 5-92.
 FT REPEAT 16884 16903 5-93.
 FT REPEAT 16925 16944 5-94.
 FT REPEAT 16966 16985 5-95.
 FT REPEAT 17007 17026 5-96.
 FT REPEAT 17048 17067 5-97.
 FT REPEAT 17089 17108 5-98.
 FT REPEAT 17130 17149 5-99.
 FT REPEAT 17171 17190 6-00.
 FT REPEAT 17212 17231 6-01.
 FT REPEAT 17253 17272 6-02.
 FT REPEAT 17294 17313 6-03.
 FT REPEAT 17335 17354 6-04.
 FT REPEAT 17376 17395 6-05.
 FT REPEAT 17417 17436 6-06.
 FT REPEAT 17458 17477 6-07.
 FT REPEAT 17499 17518 6-08.
 FT REPEAT 17540 17559 6-09.
 FT REPEAT 17581 17600 6-10.
 FT REPEAT 17622 17641 6-11.
 FT REPEAT 17663 17682 6-12.
 FT REPEAT 17704 17723 6-13.
 FT REPEAT 17745 17764 6-14.
 FT REPEAT 17786 17805 6-15.
 FT REPEAT 17827 17846 6-16.
 FT REPEAT 17868 17887 6-17.
 FT REPEAT 17909 17928 6-18.
 FT REPEAT 17950 17969 6-19.
 FT REPEAT 17991 18010 6-20.
 FT REPEAT 18032 18051 6-21.
 FT REPEAT 18073 18092 6-22.
 FT REPEAT 18114 18133 6-23.
 FT REPEAT 18155 18174 6-24.
 FT REPEAT 18196 18215 6-25.
 FT REPEAT 18237 18256 6-26.
 FT REPEAT 18278 18297 6-27.
 FT REPEAT 18319 18338 6-28.
 FT REPEAT 18360 18379 6-29.
 FT REPEAT 18401 18420 6-30.
 FT REPEAT 18442 18461 6-31.
 FT REPEAT 18483 18502 6-32.
 FT REPEAT 18524 18543 6-33.
 FT REPEAT 18565 18584 6-34.
 FT REPEAT 18606 18625 6-35.
 FT REPEAT 18647 18666 6-36.
 FT REPEAT 18688 18707 6-37.
 FT REPEAT 18729 18748 6-38.
 FT REPEAT 18770 18789 6-39.
 FT REPEAT 18811 18830 6-40.
 FT REPEAT 18852 18871 6-41.
 FT REPEAT 18893 18912 6-42.
 FT REPEAT 18934 18953 6-43.
 FT REPEAT 18975 18994 6-44.
 FT REPEAT 19016 19035 6-45.
 FT REPEAT 19057 19076 6-46.
 FT REPEAT 19098 19117 6-47.
 FT REPEAT

RA MEDLINE-20150912; PubMed-10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jaisels K., Kariyasek A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Ransdram M.A., Rutherford K.M., van Vleet A.H.M.,
 RA Whitehead S., Barrall B.G.,
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position
 CC 54 (M-5-054) in all tRNA (By similarity).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
 CC homocysteine + tRNA containing thymine.
 CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRMA
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AL39076; CAB73096.1; -
 DR InterPro: IPR000051; SAM_Dind.
 DR InterPro: IPR001566; TRMA.
 DR PROSITE: PS01230; TRMA_1; 1.
 DR PROSITE: PS01231; TRMA_2; FALSE_NEG.
 KM Transferase: Methyltransferase; tRNA processing; Complete proteome.
 FT DOMAIN 207 213 S-ADENOSYLMETHIONINE BINDING (BY
 FT SIMILARITY).
 FT ACT_SITE 315 315 BY SIMILARITY.
 SQ SEQUENCE 357 AA; 42276 MW; CEC5328347CEEA97 CRC64;
 Query Match 6.8%; Score 86; DB 1; Length 357;
 Best Local Similarity 21.8%; Pred. No. 4; Indels 62; Gaps 6;
 Matches 36; Conservative 23; Mismatches 44;
 QY 79 KRHDFKSAPEGVQORTIFPKDGNKTRAEVKE--EGDTLV----- 120
 Db 14 EKHSFIKKYKREFYTKFKFLASDKDKHRTAEISFYHENDTLFYAMFDPKSKKYLIEY 73
 QY 121 -----NRIELK--GIDPKEDGNILGKLEYNN----- 146
 Db 74 LDEADEKICAFMPRLLEYLRQDNKKELGVEFLTYTKOELSTILYHKNIEDIKSNLEN 133
 QY 147 -SHNVYIMADKOKNGIKVNFKIRH-----NIDGSOVLADYYOON 185
 Db 134 LSNILHNLIAKSKGKILFKTENTLROTLMNIDRKI---FYEFN 174
 RESULT 12
 ID NECL_RAT STANDARD; PRT; 752 AA.
 AC P28840;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 16-CCV-2001 (Rel. 40, Last annotation update)
 DE Neuroendocrine convertase 1 precursor (EC 3.4.21.93) (NCL 1) (PCL)
 DE (Prothormone convertase 1) (Proprotein convertase 1).
 GN PCSK1 OR NECL OR NEC-1 OR BDP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92168040; PubMed-1791845;
 RA Bloomquist B.T., Eipper B.A., Mains R.E.;
 RT "Prothormone-converting enzymes: regulation and evaluation of function
 RT using antisense RNA.";
 RL Mol. Endocrinol. 5:2014-2024(1991).

RA [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-92063860; PubMed-1954888;
 RA Hakes D.J., Birch N.P., Mezey A., Dixon J.E.;
 RT "Isolation of two complementary deoxyribonucleic acid clones from a
 RT rat insulinoma cell line based on similarities to Kex2 and furin
 RT sequences and the specific localization of each transcript to
 RT endocrine and neuroendocrine tissues in rats.";
 RL Endocrinology 129:3053-3063(1991).
 CC -1- FUNCTION: INVOLVED IN THE PROCESSING OF HORMONE AND OTHER PROTEIN
 CC PRECURSORS AT SITES COMPRISED OF PAIRS OF BASIC AMINO ACID
 CC RESIDUES. SUBSTRATES INCLUDE POMC, RENIN, ENKEPHALIN, DYNORPHIN,
 CC SOMATOSTATIN AND INSULIN.
 CC -1- CATALYTIC ACTIVITY: Release of protein hormones, neuropeptides and
 CC renin from their precursors, generally by cleavage of -Lys-Arg-|-
 CC bonds.
 CC -1- COFACTOR: CALCIUM DEPENDENT.
 CC -1- SUBCELLULAR LOCATION: LOCALIZED IN THE SECRETION GRANULES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M76705; AAA40945.1; -
 DR EMBL: M83745; AAA41476.1; -
 DR PIR: A41556; KXRTCL.
 DR HSP: Q45670; IDBI.
 DR MEROPS: S08.072; -
 DR InterPro: IPR002884; P_domain.
 DR InterPro: IPR000209; Peptidase_S8.
 DR Pfam: PF00082; Peptidase_S8; 1.
 DR Pfam: PF01483; P; PARTIAL.
 DR PRINTS: PR00723; SUBTILISIN.
 DR PRODOM: PD000717; P_domain; 1.
 DR PROSITE: PS00136; SUBTILASE_ASP; 1.
 DR PROSITE: PS00137; SUBTILASE_HIS; 1.
 DR PROSITE: PS00138; SUBTILASE_SER; 1.
 KM Hydrolyase; Serine protease; Glycoprotein; Zymogen; Calcium; Signal.
 FT SIGNAL 1 27 POTENTIAL.
 FT PROPEP 28 110 POTENTIAL.
 FT CHAIN 111 752 NEUROENDOCRINE CONVERTASE 1.
 FT DOMAIN 122 410 CATALYTIC.
 FT ACT_SITE 739 751 AMPHIPATHIC (POTENTIAL).
 FT ACT_SITE 167 167 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT CARBOHYD 173 173 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 401 401 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 645 645 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 366 366 T -> TT (IN REF. 2).
 FT CONFLICT 514 514 E -> A (IN REF. 2).
 SQ SEQUENCE 752 AA; 84120 MW; F630AD830A076DED CRC64;
 Query Match 6.7%; Score 85.5; DB 1; Length 752;
 Best Local Similarity 25.1%; Pred. No. 11;
 Matches 58; Conservative 25; Mismatches 77; Indels 71; Gaps 13;
 QY 17 ELDDVNGHK-----FSVSGEGDVTYGLTKLFICTGKLP-----VPMPLVTTFSYG 67
 Db 540 ERTISNGFRNMFNMGVNGENPV--GTWTLKVTDSMGQMEGRITVNMKLL-----HG 593
 QY 68 VOCFSRYPDHAKHDFKSAPEGVQORTIFPKDGNKTRAEVKEFGDTLVNRLG 127
 Db 594 T---SSQPEHMKO-----PRVYTSYNTV-----QNDRGV-----EKWNVVEEKP 631
 QY 128 IDKFEKGNILGHLEYNYNSHNVYIMADKOKNGIKVNFKIRHNIEDGSOVLADYYOONP 187
 Db 632 TONSLNGNLLVPR---NSSSSVEDRRDEOVGAPSKAMLR-----LLQSAFASKNTP 680

OY 188 ILDGPVLLPDNHYLSTQALSCKDPNEKRDHVLLEFYTAGITHGMDELYK 238
 DB 681 -----SKOS--SKIPSAKLS-----PYEGSLYALEKLNK 708

RESULT 13
 ITH3_MESAU
 ID ITH3_MESAU STANDARD: PRT: 886 AA.
 AC P97280;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Inter-alpha-trypsin inhibitor heavy chain H3 precursor (ITI heavy chain H3) (Hc3).
 GN ITH3.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Mesocricetus.
 OC Mesocricetus.
 OX NCBI_TaxID=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=97420688; PubMed=9276673;
 RA Nakatani T., Suzuki Y., Yamamoto T., Sinochana H.;
 RT "Molecular cloning and sequencing of cDNAs encoding three heavy-chain precursors of the inter-alpha-trypsin inhibitor in Syrian hamster: RT implications for the evolution of the inter-alpha-trypsin inhibitor heavy chain family.";
 RT J. Biochem. 122:71-82(1997).
 RL [2]
 RP SEQUENCE OF 31-50; 446-472 AND 504-523, AND SUBUNITS.
 RC TISSUE=Plasma;
 RX MEDLINE=97018241; PubMed=8664857;
 RA Yamamoto T., Yamamoto K., Sinochana H.;
 RT "Inter-alpha-trypsin inhibitor and its related proteins in Syrian hamster urine and plasma.";
 RT J. Biochem. 120:145-152(1996).
 RL [1]
 RP FUNCTION: MAY ACT AS A CARRIER OF HYALURONAN IN SERUM OR AS A BINDING PROTEIN BETWEEN HYALURONAN AND OTHER MATRIX PROTEIN, INCLUDING THOSE ON CELL SURFACES IN TISSUES TO REGULATE THE LOCALIZATION, SYNTHESIS AND DEGRADATION OF HYALURONAN WHICH ARE ESSENTIAL TO CELLS UNDERGOING BIOLOGICAL PROCESSES (BY SIMILARITY).
 CC -1- SUBUNIT: I-ALPHA-I PLASMA PROTEASE INHIBITORS ARE ASSEMBLED FROM ONE OR TWO HEAVY CHAINS (H1, H2 OR H3) AND ONE LIGHT CHAIN, BIKUNIN. INTER-ALPHA-INHIBITOR (I-ALPHA-I) IS COMPOSED OF H1, H2 AND BIKUNIN. INTER-ALPHA-LIKE INHIBITOR (I-ALPHA-LI) OF H2 AND BIKUNIN, AND PRE-ALPHA-INHIBITOR (P-ALPHA-I) OF H3 AND BIKUNIN.
 CC -1- PM: HEAVY CHAINS ARE INTERLINKED WITH BIKUNIN VIA A CHONDROITIN 4-SULFATE BRIDGE TO THEIR C-TERMINAL ASPARTATE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ITH FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 VWFA DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; D89287; BAA13940.1; -;
 DR InterPro: IPR002035; VWFA_A.
 DR Pfam: PF00092; vwa; 1.
 DR SMART: SM00327; VWFA; 1.
 DR PROSITE: PS50234; VWFA; 1.
 KM Serine protease inhibitor; Repeat; Signal; Multigene family;
 KM Glycoprotein.
 FT SIGNAL 1 18 POTENTIAL.
 FT PROPEP 19 30 BY SIMILARITY.

FT CHAIN 31 646 INTER-ALPHA-TRYPsin INHIBITOR HEAVY CHAIN
 FT H3.
 FT PROPEP 647 886 BY SIMILARITY.
 FT DOMAIN 279 439 VWFA.
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 577 577 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT BINDING 646 646 CHONDROITIN 4-SULFATE, CROSS-LINK SITE
 FT (BY SIMILARITY).
 SO SEQUENCE 886 AA; 99018 MW; AC0594C685257688 CXC64;
 Query Match 6.7%; Score 85.5; DB 1; Length 886;
 Best Local Similarity 23.8%; Pred. No. 13;
 Matches 36; Conservative 34; Mismatches 62; Indels 19; Gaps 7;

OY 74 YPDHMKRDPFKSMPEGYVQRT-----IFKDDGNYKTRAEVKFGSD---TLVNRIE 124
 DB 476 YPENAIL-DLTKNSYPHYDGEFAVAGRLADSDMMNFK--ADVKGHALNDLFTFEVD 532
 OY 125 LKSID--FKEDGNLGHKLEYNSHNYIMADKONGIKYFKIRHNIEDGSVQLADY 182
 DB 533 MKEDDALEKGGYIFGNTIERMAYLVTEQLERKN--AHGEKENLQAALSLUKY 589
 OY 183 QANTPIIDGPVLLPDNHYLSTQALSCKDPNE 213
 DB 590 HFTVPLPMPVYTKPEDN--EDQTSIDKPGE 618

RESULT 14
 TRXB-CHLTR
 ID TRXB-CHLTR STANDARD: PRT: 312 AA.
 AC 084101;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thioedoxin reductase (EC 1.6.4.5) (TRXB).
 GN TRXB OR CT099.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UM-3/Cx;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis.";
 RT Science 282:754-759(1998).
 CC -1- CATALYTIC ACTIVITY: NADPH + oxidized thioedoxin = NADP(+) + reduced thioedoxin.
 CC -1- COFACTOR: FAD.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
 CC -1- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
 CC -1- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE OXIDOREDUCTASES CLASS-II.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; AE001284; AAC67690.1; ALT_INIT.
 DR HSSP: Q39243; 1YDC.
 DR PHCI-2DPAGE; 084101; -;
 DR InterPro: IPR00759; Adrndx_reductase.
 DR InterPro: IPR001327; FAD_pyr_redox.
 DR InterPro: IPR001100; Pyr_redox.
 DR InterPro: IPR000103; Pyridine_redox_2;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:54:24 ; Search time 40 Seconds
(without alignments)
572.000 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGEELFTAVVPILVELDG.....VLEFVTAAGITGHGDELTK 238

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: PIR_73:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136	97.3	238	1 J01514	green-fluorescent
2	105.5	8.3	2	H72228	hypothetical prote
3	100.5	7.9	887	2 E82590	leucyl-tRNA synthet
4	99	7.8	336	2 C64468	hypothetical prote
5	92.5	7.3	2573	2 D71614	hypothetical prote
6	90	7.1	471	2 T27856	hypothetical prote
7	90	7.1	797	2 JC4078	protective surface
8	90	7.1	808	2 F64102	protective surface
9	89.5	7.0	393	2 C64613	conserved hypotet
10	88.5	7.0	312	2 C81710	thioredoxin reduct
11	88.5	7.0	425	2 C97354	hypothetical prote
12	87.5	6.9	861	2 H64102	leucine-tRNA ligas
13	87	6.9	822	2 C71633	ATP-dependent nucl
14	87	6.9	2334	2 S32920	cell wall-associat
15	86	6.8	357	2 G81355	tRNA (uracil-5-)-m
16	85.5	6.7	752	2 KAKTCL	proprotein convert
17	85.5	6.7	836	1 J0V1DL	DNA-directed DNA p
18	85.5	6.7	889	2 JC5576	inter-alpha-trypsi
19	84.5	6.7	351	2 B71556	probable thioredox
20	84.5	6.7	860	2 AC0582	leucyl-tRNA synthet
21	83	6.5	281	2 AD2052	hypothetical prote
22	83	6.5	533	2 S47271	AHNK-related prot
23	83	6.5	1092	2 S42798	fibronectin-bindin
24	83	6.5	3472	2 T31308	hypothetical 367K
25	82.5	6.5	613	2 A99552	oligodeoxyribosidase
26	82.5	6.5	1134	2 A60234	Iga Fc receptor pr
27	82.5	6.5	1164	1 FCSOAG	Iga Fc receptor pr
28	82	6.5	207	2 C70483	hypothetical prote
29	82	6.5	653	1 A11535	fructose-bisphosph

30	82	6.5	764	2 T25012	hypothetical prote
31	82	6.5	1082	2 T41988	hypothetical prote
32	82	6.5	1433	2 B83952	DNA polymerase III
33	82	6.5	2166	2 G70163	hypothetical prote
34	81	6.4	647	2 H89988	hypothetical prote
35	81	6.4	953	2 S5156	probable membrane
36	80.5	6.3	355	1 LKCH	ubiquinol-cytochr
37	80.5	6.3	504	1 S23558	hypothetical prote
38	80.5	6.3	636	2 C70031	hypothetical prote
39	80.5	6.3	655	2 D83917	DNA topoisomerase
40	80.5	6.3	1334	2 AB1775	hypothetical prote
41	80.5	6.3	1346	2 G71613	hypothetical prote
42	80.5	6.3	2222	1 A36028	DNA-directed DNA p
43	80.5	6.3	6359	2 T31679	bactinacin synthet
44	80	6.3	311	2 C97349	uncharacterized pr
45	80	6.3	393	2 E82897	hypothetical prote

ALIGNMENTS

RESULT 1

J01514

green-fluorescent protein [validated] - hydromedusa (Aequorea victoria)

C:Species: Aequorea victoria

C:Date: 03-Dec-1999 #sequence-revision 03-Dec-1999 #text-change 23-Mar-2001

C:Accession: J01514; P00335; S48693; S51330; S51331

R:Prasner, D.C.; Eckenrode, V.K.; Ward, W.W.; Prendergast, F.G.; Cormier, M.J.

Gene 111, 229-233, 1992

A:Title: Primary structure of the Aequorea victoria green-fluorescent protein.

A:Reference number: J01514; MUID:92175527; PMID:1347277

A:Accession: J01514

A:Molecule type: DNA

A:Residues: 1-107,'S',109-238 <PRA1>

A:Cross-references: GB:M62654; NID:9155662; PIDN:AAA27722.1; PID:9155663

A:Accession: J01514

A:Molecule type: mRNA

A:Residues: 1-99,'F',101-140,'L',142-218,'V',220-238 <PRA2>

A:Cross-references: GB:M62653; NID:9155660; PIDN:AAA27721.1; PID:9155661

A:Accession: P00335

A:Molecule type: Protein

A:Residues: 46-64;74-122;132-151;154-183;185-200 <PRA3>

R:Inouye, S.; Tsuji, F.I.

FEBS Lett. 351, 211-214, 1994

A:Title: Evidence for redox forms of the Aequorea green fluorescent protein.

A:Reference number: S48693; MUID:94364470; PMID:8082767

A:Accession: S48693

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-156,'P',158-171,'K',173-238 <INO>

A:Cross-references: GB:I29345; NID:9606383; PIDN:AAA58246.1; PID:9606384

R:Watkins, J.N.; Campbell, A.K.

Submitted to the EMBL Data Library, January 1995

A:Reference number: S51330

A:Molecule type: mRNA

A:Residues: 1-13,'V',15-24,'Q',26-44,'N',46-153,'G',155-156,'P',158-171,'K',173-227,'

A:Cross-references: EMBL:X83959; NID:9634008; PIDN:CAA58789.1; PID:9634009

A:Experimental source: clone gtfp

A:Accession: S51331

A:Molecule type: mRNA

A:Residues: 1-24,'Q',26-29,'R',31-83,'L',85-153,'G',155-156,'P',158-171,'K',173-208,'

A:Cross-references: EMBL:X83960; NID:9634010; PIDN:CAA58790.1; PID:9634011

A:Experimental source: clone gtfp2

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Submitted to the Brookhaven Protein Data Bank, August 1996

A:Reference number: A65692; PDB:1GFL

A:Contents: annotation: X-ray crystallography, 1.9 angstroms, residues 'A',2-79,'R',8

A:Note: engineered sequence based on J01514, cloned and expressed in Escherichia coli

R:Yang, F.; Moss, L.G.; Phillips Jr., G.N.

Nat. Biotechnol. 14, 1246-1251, 1996

A:Title: The molecular structure of green fluorescent protein.

A:Reference number: A58953; MUID:98294543; PMID:9631087

A:Contents: annotation; X-ray crystallography, 1.9 angstroms
 C:Comment: This protein is excited by the photoprotein aequorin (see PIR:AQJFNV) emitting
 C:Comment: The chromophore of this protein is formed by modification of Ser-dehydro-Tyr-
 C:Genetics:
 A:Gene: GFP
 A:Introns: 69/3; 167/3
 C:Superfamily: green-fluorescent protein
 C:Keywords: chromoprotein; luminescence
 F:65-67/Cross-link: 5-imidazolone (Ser-Gly) #status experimental
 F:66/Modified site: dehydrotyrosine (Tyr) #status experimental

Query Match 97.3%; Score 1236; DB 1; Length 238;
 Best Local Similarity 96.2%; Pred. No. 5.2e-96;
 Matches 229; Conservative 6; Mismatches 3; Indels 0; Gaps 0;

QY 1 MSKGELEFVAVPLVLELDGVDNGHKSFSVSGEGSDVYGLTKLFICTTGKLPVPTL 60
 DB 1 MSKGELEFVAVPLVLELDGVDNGHKSFSVSGEGSDVYGLTKLFICTTGKLPVPTL 60
 QY 61 VTTFSSVQCFSRYPDHRKRDFFKSAMPEGVQORTIFPKDGNKTRAEVKEGDTLV 120
 DB 61 VTTFSSVQCFSRYPDHRKRDFFKSAMPEGVQORTIFPKDGNKTRAEVKEGDTLV 120
 QY 121 NRIELKIDFEKDSNIIIGHKLEYNSHNYIMADKDKNGKVKFKIRHNEDGSVOLAD 180
 DB 121 NRIELKIDFEKDSNIIIGHKLEYNSHNYIMADKDKNGKVKFKIRHNEDGSVOLAD 180
 QY 121 NRIELKIDFEKDSNIIIGHKLEYNSHNYIMADKDKNGKVKFKIRHNEDGSVOLAD 180
 DB 121 NRIELKIDFEKDSNIIIGHKLEYNSHNYIMADKDKNGKVKFKIRHNEDGSVOLAD 180
 QY 181 YVQONTPILDGVPVLPNNHLYSTOSALSKDPNEKRDMVLEFVTAAGITGHMDELK 238
 DB 181 YVQONTPILDGVPVLPNNHLYSTOSALSKDPNEKRDMVLEFVTAAGITGHMDELK 238

RESULT 2

H72228
 hypochlorite protein TM1624 - Thermotoga maritima (strain MSB8)

C:Species: Thermotoga maritima
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
 C:Accession: H72228
 R:Neelson, K.E.; Clayton, R.A.; Gill, S.R.; Dodson, R.J.; Haft, D.H.; Hickey
 Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.;
 C.M.
 Nature 399, 323-329, 1999
 A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
 A:Reference number: A72200; MUID:99287316; PMID:10360571
 A:Accession: H72228
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-785 <ARN>
 A:Cross-references: GB:AE001806; GB:AE00512; NID:94982196; PIDN:ADJ36691.1; PID:9498219
 A:Experimental source: strain MSB8
 C:Genetics:
 A:Gene: TM1624

Query Match 8.3%; Score 105.5; DB 2; Length 785;
 Best Local Similarity 23.9%; Pred. No. 0.81;
 Matches 43; Conservative 25; Mismatches 65; Indels 47; Gaps 7;

QY 21 DVNCH-----KFSVSGEGSDVYGLTKLFICTTGKLPVPTLVTFSSVQCF 72
 DB 5 DLNFGVSVRDNKGFSEFPGV-VQADLVK-----GLLPHRYGM----- 46
 QY 73 RYPDHMRHDFEKSAMPEGVQORTIFPKDGNKTRAEVKEGDTLVNRIELKIDFKE 132
 DB 47 -----NEDLFKIEEDREWIYEREFKEFKEGGEVDLVFEVDLSDVLLNGVLA- 97
 QY 133 DGNILGHKLLEYNSHNYIMADKDKNGKVKFKIRHNEDGSVOLADYVQONTPIIDP 192
 DB 98 -GSTEDMFTIRPDVTVL-----KKNHLKVIK-----SPIRVEKTLQONTGVLAGP 145

RESULT 3

E82590
 leucyl-tRNA synthetase XF2176 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: E82590
 R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717; PMID:10910347
 A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: E82590
 A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-887 <STM>
 A:Cross-references: GB:AE004031; GB:AE003849; NID:99107309; PIDN:AAF84975.1; GSPDB:GN
 A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Dorri, H.; Feliciani, A.P.; Ferreira, A.J.S.
 submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sava
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
 M.; Tsunako, M.H.; Vailidis, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328
 A:Contents: annotation
 C:Genetics:
 A:Gene: XF2176
 C:Superfamily: leucine-tRNA ligase

Query Match 7.9%; Score 100.5; DB 2; Length 887;
 Best Local Similarity 23.2%; Pred. No. 2.5;
 Matches 46; Conservative 28; Mismatches 67; Indels 57; Gaps 10;

QY 49 TTGKLPVPTLVTFSSVQCFSRYPDHRKRDFFKSAMPEGVQORTIFPKDGNV-- 106
 DB 329 TNEQLPV-WVANFVLMVGGAVAVPGHQRDEF--ARKYGLPIRQVIALKEPKNDE 385
 QY 107 -----KTRAEVKEGDTLVNRIELKIDFKEKDGNIIGHKLEYNSHNYI 152
 DB 386 STWEPDVRWRYADKTR---EFE---LINAEPGLDQAFVLAERF----- 429
 QY 153 MADKDKNG-IVNFKIRHNEDGSVOLADYVQONTPII-----DGPVLLPDN 198
 DB 430 ---RGRGQRVNVYRLR---DMGVSRRYWCPCIPVYCPGCAVPVPEQJLPVILPEN 482
 QY 199 -HYISTOSALSKDPNEKR 215
 DB 483 VAFSGTSPKTPDEWKR 500

RESULT 4

C64468
 hypochlorite protein MJ1348 - Methanococcus jannaschii

C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
 C:Accession: C64468
 R:Bolt, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 ; Reich, C.I.; Overbeek, R.; Kirsch, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
 rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996

A:Title: Complete genome sequence of the methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999; PMID:8688087
 A:Accession: C64468
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-336 <BDL>
 A:Cross-references: GB:U67574; GB:L77117; NID:91591978; PIDN:AAB99360.1; PID:91591990
 C:Genetics:
 A:Map position: FOR1295121-1296131

A:Start codon: TTG

Query Match 7.8%; Score 99; DB 2; Length 336;
Best Local Similarity 24.3%; Pred. No. 0.93;
Matches 67; Conservative 40; Mismatches 65; Indels 104; Gaps 19;

OY 19 DGDVNGHKSFSVGE-----GEGDVYTGKLTFLICTYTGKLPYV-----57
DB 91 DGDY-----YNSGELSTIASIFAKIGKLDI-----TKNPLMSGGE-----WIINDYKDN 137
OY 58 -----PLVYTFST--GVQCFSRYPDMKR-----HDFKSMPEGYOQRTIF 99
DB 138 SEDILKSVLTDFDSYEERKEILNREP--HLRKLFDNIYFNNSDFEDFMK-----MF 187
OY 100 FKDGNYKTRADEVKE-----GDTLVNRIELKGDPEKDGNIIL--GHKLEYNYS 147
DB 188 FIGAGNMRKFLFEEVEEFENKIKSCQISNEYNEI--IKRPD--KMSDLAIALHLKENEK 245
OY 148 HNYIYIMADK-----ORNGIKVNFKIRHNIEDGSV-----OLADYVOQNTPI 189
DB 246 CLYYVMEKEFEDDEFEIEKKNINLTVDIAYNKNNGVKKKEWLNLENYKE---II 302
OY 190 DGPVILPDNHYLSTQSALSKD--PNEKRDMVLEFV 224
DB 303 KRP--LPNTY-----KDAHNDLNLIELDYVYLKEFI 332

RESULT 5

hypothetical protein PF0460c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: D71614
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Perlea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.;
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: D71614
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2573 <GAR>
A:Cross-references: GB:AE001396; GB:AE001362; NID:9384518; PIDN:AACT1881.1; PID:9384519
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0460c

Query Match 7.3%; Score 92.5; DB 2; Length 2573;
Best Local Similarity 26.2%; Pred. No. 4.7;
Matches 34; Conservative 31; Mismatches 52; Indels 13; Gaps 5;

OY 93 VOQRTIFPKD--DGKYKTRAEYKFEGLVLRNIELKIGDKEDGMILGHKLEYN--YNSH 148
DB 126 LKKEITLCKDKSGNDPMDSETSLFKDMDVDKEIK--DFESSIKIKKREYNYNTYNNK 183
OY 149 NYIYIMADKONGIKYKIRHNIEDGSYQADYVOQNTPIIDGPVILPDNHYLSTQSALS 208
DB 184 NLIHIEKKNKKDEKKNKHNNDENNMM---IYYKN---IDKTYIIDNNVYHILINDIN 236
OY 209 KDPNEKRDM 218
DB 237 TYLKREDDYM 246

RESULT 6

hypothetical protein ZK418.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T27856
R:Fulton, L.
Submitted to the EMBL Data Library, April 1994
A:Description: The sequence of C. elegans cosmid ZK418.

A:Reference number: Z20430

A:Accession: T27856
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-471 <FUL>
A:Cross-references: EMBL:000047; PIDN:AAA50666.1; CESP:ZK418.2
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:ZK418.2
A:Introns: 41/3; 59/1; 139/1; 214/3; 241/2; 295/3; 329/3; 381/3; 454/2
C:Superfamily: Caenorhabditis elegans hypothetical protein ZK418.2

Query Match 7.1%; Score 90; DB 2; Length 471;

Best Local Similarity 23.3%; Pred. No. 8.2;
Matches 49; Conservative 34; Mismatches 81; Indels 46; Gaps 10;

OY 46 FICTGKLPVPPVPLVTFSTYGVQCFSRYPDMKRHDF--FKSAMP-----GYVOQRTI 98
DB 140 FLCKFSVLPV-----SHTY-----RYPIEMKKGVPVTFKPEPELINDVKKFTVNNEM 186
OY 99 FF-----KDD--GNKTRAEYKE-----GDTLVNRIELKGDPEKDGNIILGHKLEYN 146
DB 187 FFDLIKIDNGSYSCDTLVTLDSFKTMTNGVAIINVMGTIGEFYIGNVYVNLKEAHV 246
OY 147 SHNYIYIMADKONGIKY--NFKIRHNIEDGSYQADYVOQNTPIIDGPVIL-----195
DB 247 VPNSYKIIDKIFTEGKVPQCDWVFYVKNQSP--SDPHVHGVLADSSIVMAAAYVSHMSI 304
OY 196 -PDNHYLSTQSALSKDPEKRDMVLEFV 224
DB 305 GFOIAMISHENLIFNLTKSRDMHILTYI 334

RESULT 7

protective surface antigen D-15 precursor - Haemophilus influenzae (type b)
C:Species: Haemophilus influenzae
A:Variety: type b
C:Date: 30-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 29-Sep-1999
C:Accession: J04078
R:Flack, F.S.; Loosmore, S.; Chong, P.; Thomas, W.R.
Gene 156, 97-99, 1995
A:Title: The sequencing of the 80-kDa D15 protective surface antigen of Haemophilus 1
A:Reference number: J04078; MUID:95255676; PMID:7737523
A:Accession: J04078
A:Molecule type: DNA
A:Residues: 1-797 <FLA>
A:Cross-references: GB:U13961; NID:9537447; PIDN:AAA85645.1; PID:9537448
A:Experimental source: type b
C:Superfamily: protective surface antigen D-15
C:Keywords: surface antigen
F:1-19/DNA: signal sequence #status predicted <SIG>
F:20-797/Product: protective surface antigen D-15 #status predicted <MAT>

Query Match 7.1%; Score 90; DB 2; Length 797;

Best Local Similarity 20.2%; Pred. No. 16;
Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

OY 22 VNGHK-----FVSGBGSDVTYKLIKIFCTTGKLPVPPVPLVTFSTYGVQCFSRYP 75
DB 402 INGSNDEVDVYKVKERTGSIKNG-----IGYTESGISYO 438
OY 76 DHMKRHDF-----FKSAMPEGYVOQ-----RTIEFKDGGVYKTRA 110
DB 439 ASVKODNPLGCGAASVIGTKNDYGVSLGTEPYFTKQGVSLGQNVFFENYDMSKST 498
OY 111 EYKFGDTLVNRIELKIGDKFEDGNI--LGH-----KLEYNNSHNYIYIMADKOK- 158
DB 499 SGNRYKRTYGSVWL--GPPVNNNSYVYGLGTYKIKISNFALEYN--RNLYIQSMKFG 554
OY 159 NGIKYV--KIRIRHNIEDGSYQADYVOQNTPI--LDGPVILP--DNHYLSTQSALSADPNEK 214
DB 555 NGIKTNDEDFSGWYNSLNRGYPPTKGVKASLAGRVITPGSDNRYKLSADVOGFFPLD 614

QY 215 RDHMLLEFVTAAGITHG 232
 Db 615 RDHMLVVSASAKSAGYANG 632

RESULT 8

F64102
 C:Species: Haemophilus influenzae
 C:Date: 13-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Sep-1998
 C:Accession: F64102
 R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.; Science 269, 496-512, 1995
 A:Authors: Guelm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: F64102
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-808 <TIGR>
 A:Cross-references: GB:AE000587; TIGR:HT0917
 C:Superfamily: protective surface antigen D-15
 C:Keywords: surface antigen

Query Match 7.1%; Score 90; DB 2; Length 808;
 Best Local Similarity 20.2%; Pred. No. 17;
 Matches 52; Conservative 36; Mismatches 96; Indels 74; Gaps 12;

QY 22 VNGHK-----FVSQSGEGDYVYKLTFLFCTTGKLPVWPVLTFTSYGOCESRRYP 75
 Db 415 INGSNDEVDVYVYKERNKTSING-----IGTSGSISYQ 451
 QY 76 DHMKRHF-----FKSAMPEGVYQ-----RTIFFDQCNVYTRA 110
 Db 452 ASYKQDNFLGTGAANVSTAGTKNDYGTGVNIGTPTTKQCVSLGCVNFFENYNSKSDT 511
 QY 111 EVFEGDTLVNRIELKIDFKEDGNI---LGH-----KLEYNYSNHYVIMADKQK 158
 Db 512 SSNKRRTYGSNVTL-GFPVNNNSYVYGHTYKNSNPALEVN---RNLYIGSMKFKG 567
 QY 159 NGIKVN-FKIRHNEDSSVOLADYVQONTPL-LDGPVLLP--DNHYISTOSALSKDENEK 214
 Db 568 NGIKTNDGDFSGMNNYSLNKGYFPTRKVASLGGRTVIGSDMKYKLSADYGFYPLD 627
 QY 215 RDHMLLEFVTAAGITHG 232
 Db 628 RDHMLVVSASAKSAGYANG 645

RESULT 9

C64613
 C:Species: Helicobacter pylori
 C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
 C:Accession: C64613
 R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalaf, H.G.; Glodek, A.; McKenney, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Matthey, L.; Nature 388, 539-547, 1997
 A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.; Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
 A:Reference number: A64520; MUID:97394467; PMID:9252185
 A:Accession: C64613
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-393 <TOM>

Query Match 7.0%; Score 89.5; DB 2; Length 393;
 Best Local Similarity 20.8%; Pred. No. 7.1;

Matches 51; Conservative 42; Mismatches 77; Indels 75; Gaps 15;

QY 34 EGDVYTKLTLK-----FICTGKLPVWPVLTFTSYGOCESRRYDMKRH-- 81
 Db 174 QGD---GRVLESMPNHRCEKIFV---HPPVPM-----NEKKHR 207
 QY 82 ----DFKSAW---PEGVVOQRT---IFEKD-----GNKTRAEVKFEGD-TLVRIE 124
 Db 208 VLSEKFLNEALRYLKPFGFELRTFDDSLYFEDSLKALKNFQCEIEIKKNAQIPVSKYE 267
 QY 125 LKSIDFKED-GNIGHKLEYN---YNSHNYIMADK---OKNGIKVNFKIRHNEDEG-SV 176
 Db 268 ARMKKLKDDIDLYLITSLMNEPFDNADPESDFTTTSKSVGTLTKTKIIQEGYFV 327
 QY 177 QLADYVOQNTPL-----DGP-----VLLPDNHYISTOSALSKDENEKRDHMLLEFV 224
 Db 328 HVCNITENKQGFVLELSMGSDPDMVRLFVLTENQIYLNKSPKLTNNKHAHLLONIL 387
 QY 225 TAAGI 229
 Db 388 SOKGI 392

RESULT 10

C81710
 C:Species: Chlamydia muridarum, Chlamydia trachomatis Mopn
 C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000
 C:Accession: C81710
 R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, Nucleic Acids Res. 28, 1397-1406, 2000
 A:Title: Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39
 A:Reference number: A81500; MUID:20150255; PMID:10684935
 A:Accession: C81710
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-312 <TEFT>
 A:Cross-references: GB:AE002304; GB:AE002160; NID:97190409; PIDN:AAF39233.1; PID:9719
 A:Genetic source: strain N1gg (MOPn)
 C:Genetics: C:Gene: TC0375
 C:Superfamily: thioredoxin reductase; thioredoxin reductase homology

Query Match 7.0%; Score 88.5; DB 2; Length 312;
 Best Local Similarity 21.1%; Pred. No. 6.4;
 Matches 48; Conservative 29; Mismatches 81; Indels 69; Gaps 9;

QY 1 MSKGEELFVAVPLVLELDGVDNGHKFSVSGEGSDVYTKLTKFICTTGKLPVWPPTL 60
 Db 97 LKSGEDFTFCDACIAT--GASAKRLISIPGAGDNFEMOKGVYACAVC-DGASPI----- 147
 QY 61 VTFESYGVQCESRYPDMKRHDF-----KSAMPEGVVOQRTIFFKDDGNVYKTRAEVKE 115
 Db 148 -----FRDRDLFVJGGGSALEAMFLR-----YGRVYVHR 161
 QY 116 GDTL-----VNRTELKIDFKEDGNIGKLEYNYSNHYVIMADKQNGIKVNEKIRH 169
 Db 182 RDTLRASKAVWNAQA-----ANKKIYFIAMSEVVKILGSLVRSI-----DIFN 225
 QY 170 NIEDGSVQL-----ADYVQONTPLIDGVYLLPDNHYISTOSALS 209
 Db 226 NVEKTVTMEAGVFFAIGHQPTAFVIGCOLSIDENGYITTERGSSR 272

RESULT 11

C97354
 C:Species: Clostridium acetobutylicum
 C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
 C:Accession: C97354
 R:Rolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, O.; Gibson, R.; Daily, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001
 A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium C16
 A:Reference number: A96900; MUID:21359325; PMID:21359325
 A:Accession: C97354
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-425 <KUR>
 A:Cross-references: GB:AE001437; PIDN:AAK81622.1; PID:G15026806; GSPDB:GN00168
 A:Experimental source: Clostridium acetobutylicum ATCC824
 C:Genetics:
 A:Gene: CAC3702

Query Match 7.0%; Score 88.5; DB 2; Length 425;
 Best Local Similarity 24.3%; Pred. No. 9.6;
 Matches 59; Conservative 29; Mismatches 82; Indels 73; Gaps 14;

QY 20 GDVNGHKEFSVSGEGEDVYTKLTKFCITTK-----LP-----VPM 57
 DB 109 GFVNGKLLPASGEAIKD--FGKV---LKNKGDKLKLTRQVEVEMFGVGRVSGPNIIEF 162
 QY 58 PLVLTFTSYGVQCSRRPDHKKRDFKKSAMPEGVQQRITFFDDGNYKTR-----AEV 112
 DB 163 SEVVKRNKNGVDI-----KDNFVAKVSEKGVKGNP---SGEGYSTEEWYNYLKE 211
 QY 113 KFEED-----TLVNRLELKGIDPKE--DGNILGHKLEYNY-----NSH- 148
 DB 212 KYGDNDNYLCTDEKSLNSSEKIDSRKRNKNGVYG-KTAVAYGENLDVYSMTSNSHO 270
 QY 149 -NVYIMADKOKNGIKRVFKIRHNIEDSGVOLADYVQONTPLDGPVLLPDNHYLSTOSAL 207
 DB 271 YNPILIDGKEFKDNMYNKLKEYGDESVAYSKYV-EHTLIEDYIKTLPDEYF--TSKGL 327
 QY 208 SKD 210
 DB 328 VKD 330

RESULT 12

H64102
 leucine-tRNA ligase (EC 6.1.1.4) - Haemophilus influenzae (strain Rd KW20)
 N:Alternate names: leucyl-tRNA synthetase
 C:Species: Haemophilus influenzae
 C>Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
 C:Accession: H64102
 R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
 ; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
 ; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630; PMID:7542800
 A:Accession: H64102
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-861 <TIGR>
 A:Cross-references: GB:U32774; GB:LA2023; NID:g1573942; PIDN:AAK22581.1; PID:g1573943; T
 C:Genetics:
 A:Gene: leus
 C:Superfamily: leucine-tRNA ligase
 C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 6.9%; Score 87.5; DB 2; Length 861;
 Best Local Similarity 24.3%; Pred. No. 29;
 Matches 45; Conservative 25; Mismatches 68; Indels 47; Gaps 9;

QY 49 TTGKLVPWPLVLTFTSYGVQCSRRPDHKKRDFKKSAMPEGVQQRITFFKDDGKNT 108
 DB 314 TGDRLPI-VVAVNFYLMHGTGAVAAVPAHQDQD--FEPAQYSLPIKQVIAPLADDEIDL 370
 QY 109 RAEVKEFGDLYVNIIELEKIDFKEDGNILGHKLEYNYNSHNYIMADK-QKNGT---KVN 164
 DB 371 TKQAFVEHGKLVNSDEDEKGNF--DGAENG-----IADLLEIKLVGAKQVNV 414

QY 165 FKIRH-----NIEDGSVOLADYVQONTPLDGPVLLPDNHYL-STOSAL 207
 DB 415 YRLDMDGVSKQRWGAFLPMLTENGDVPA-----PHEDDLPIILPEVDYMDGVASPI 467
 QY 208 SKDPN 212
 DB 468 NADPN 472

RESULT 13

C71633
 ATP-dependent nuclease chain A (addA) RP734 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C>Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: C71633
 R:Anderson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A71630; MUID:99039499; PMID:9823893
 A:Accession: C71633
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-822 <AND>
 A:Cross-references: GB:AJ235273; GB:AJ235269; NID:93861237; PIDN:CAA15163.1; PID:e134
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: addA; RP734

Query Match 6.9%; Score 87; DB 2; Length 822;
 Best Local Similarity 25.2%; Pred. No. 30;
 Matches 54; Conservative 28; Mismatches 54; Indels 78; Gaps 13;

QY 57 WPLVLTFTSYGVQCSRRPDHKKRDFKKSAMPEGVQQRITFFKDDGNTYRAEVFEG 116
 DB 473 WPLLT-----NQKQKEFFWT-LPE-----DDKNAKSADLTL-- 503
 QY 117 DTLVN-----RIELKGI-----DFKEDGNILGHKLEYNYNSHNYIMADKOKNGIKVNF 165
 DB 504 DKIVNFIKEKIKSGILSTSRSEKDFMLVAKRD--KFSN--LIKELSKAKLVEI 559
 QY 166 KIRHNIEDGSVOLADYVQONTPLD-----GPVLLPDNHYLSTOSALSKDP----- 211
 DB 560 SDRINL-----KENPIIDLLAAKFFVLLPDDDL--NLACLKSPILIGISEQKL 606
 QY 212 -----NEKRDHVVLEFVTAAGITGMD---ELKX 238
 DB 607 YTLVKKNDHLEWVLSHNDYHKLDSTIETK 640

RESULT 14

S32920
 cell wall-associated protein precursor wapa [similarity] - Bacillus subtilis
 C:Species: Bacillus subtilis
 C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 21-Jul-2000
 C:Accession: S32920; E69730; T47101
 R:Poster, S.J.
 Mol. Microbiol. 8, 299-310, 1993
 A:Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
 protein.
 A:Reference number: S32919; MUID:93302506; PMID:8316082
 A:Accession: S32920
 A:Molecule type: DNA
 A:Residues: 1-2334 <ROS>
 A:Cross-references: GB:L05634; NID:g304177; PIDN:AAA22883.1; PID:g304179
 R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Aevedo, V.; Ber
 C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,
 Nature 390, 249-256, 1997
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M
 koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino
 A:Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau
 Y. M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 17:43:48 ; Search time 72 Seconds
(without alignments)
440.467 Million cell updates/sec

Title: US-10-071-976-2

Perfect score: 1270

Sequence: 1 MSKGELEFTAVVPILVELDG.....VLEFVTAGITGHGMDLELYK 238

Scoring table:

BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A:Geneseq_101002:*

- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1270	100.0	238	19	AAW52313
2	1267	99.8	238	19	AAW52328
3	1267	99.8	238	19	AAW52330
4	1266	99.7	238	19	AAW52333
5	1266	99.7	238	19	AAW52335
6	1265	99.6	238	19	AAW52351
7	1265	99.6	238	19	AAW52335
8	1265	99.6	238	19	AAW52336
9	1265	99.6	238	19	AAW52337
10	1265	99.6	238	19	AAW52338

11	1265	99.6	238	19	AAW52344	Engineered green f
12	1265	99.6	238	19	AAW52346	Engineered green f
13	1265	99.6	238	19	AAW52349	Engineered green f
14	1265	99.6	238	19	AAW52324	Engineered green f
15	1265	99.6	238	19	AAW52331	Engineered green f
16	1265	99.6	238	19	AAW52334	Engineered green f
17	1265	99.6	238	19	AAW52314	Engineered green f
18	1264	99.5	238	19	AAW52339	Engineered green f
19	1264	99.5	238	19	AAW52340	Engineered green f
20	1264	99.5	238	19	AAW52347	Engineered green f
21	1264	99.5	238	19	AAW52350	Engineered green f
22	1264	99.5	238	19	AAW52332	Engineered green f
23	1263	99.4	238	19	AAW52341	Engineered green f
24	1263	99.4	238	19	AAW52345	Engineered green f
25	1263	99.4	238	19	AAW52348	Engineered green f
26	1262	99.4	238	19	AAW52342	Engineered green f
27	1262	99.4	238	19	AAW52323	Engineered green f
28	1261	99.3	238	19	AAW52343	Engineered green f
29	1261	99.3	238	19	AAW52325	Engineered green f
30	1260	99.2	238	19	AAW52316	Engineered green f
31	1260	99.2	238	19	AAW52317	Engineered green f
32	1257	99.0	238	19	AAW52329	Engineered green f
33	1256	98.9	238	19	AAW52321	Engineered green f
34	1256	98.9	238	19	AAW52322	Engineered green f
35	1253	98.7	238	19	AAW52319	Engineered green f
36	1250	98.4	238	19	AAW52318	Engineered green f
37	1249	98.3	238	19	AAW52320	Engineered green f
38	1246	98.1	238	17	AAW05304	Green fluorescent
39	1246	98.1	238	18	AAW24232	Aequorea victoria
40	1246	98.1	238	19	AAW76371	A. victoria green
41	1246	98.1	238	19	AAW76105	A. victoria green
42	1246	98.1	238	19	AAW40479	Wild-type green fl
43	1246	98.1	238	22	AAE73552	Aequorea victoria
44	1246	98.1	238	19	AAE16038	A. victoria green
45	1245	98.0	238	19	AAW65084	

ALIGNMENTS

RESULT 1
AAW52313
ID AAW52313 standard; Protein; 238 AA.
XX
XX AAW52313;
XX
DT 10-JUL-1998 (first entry)
XX
DE Aequorea green fluorescent protein.
XX
AC Green fluorescent protein; engineered fluorescent protein;
KW reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Aequorea victoria.
XX
PN WO9806737-A1.
XX
PD 19-FEB-1998.
XX
PF 15-AUG-1997; 97WO-US14593.
XX
PR 30-AUG-1996; 96US-0706408.
XX
PR 16-AUG-1996; 96US-0024050.
XX
PA (AURO-) AURORA BIOSCIENCES.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PA (UOR-) UNIV OREGON.
XX
PI Cubitt AB, Heim R, Ormo WF, Remington JS, Tsien RY;
DR WPI; 1998-159454/14.
DR N-PSDB; AAV19946.

XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS Claim 1; Page 53; 120pp; English.
 XX
 CC This sequence is the wild type Aequorea victoria green fluorescent
 CC protein. The nucleic acid was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H, Y,
 CC W or F) and having fluorescent properties different from GFP. (II) are
 CC useful as reporter molecules in immunological or hybridisation assays,
 CC for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX Sequence 238 AA:
 SO
 Query Match 100.0%; Score 1270; DB 19; Length 238;
 Best Local Similarity 100.0%; Pred. No. 2e-123;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVNPILVELDGVNGHKRSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKRSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
 QY VTFPSYGVQCFSRYPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVFEEDTLV 120
 DB VTFPSYGVQCFSRYPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVFEEDTLV 120
 QY 121 NRLEKIGIDKEGDNILGHKLEYNYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLEKIGIDKEGDNILGHKLEYNYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOQNPILDGPVLLPDPNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YVOQNPILDGPVLLPDPNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 RESULT 2
 AAM52328
 ID AAM52328 standard; Protein; 238 AA.
 XX
 AC AAM52328;
 XX
 DT 10-JUL-1998 (first entry)
 XX
 DE Engineered green fluorescent protein S65T.
 XX
 KW Green fluorescent protein; engineered fluorescent protein; mutain;
 KM reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.
 OS Synthetic.
 OS Aequorea victoria.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 65 /label= S65T
 FT /note= "Ser to Thr mutation"
 FT
 XX
 XX W09806737-A1.
 PN
 XX
 PD 19-FEB-1998.

XX
 PF 15-AUG-1997; 97WO-US14593.
 XX
 PR 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIT CALIFORNIA.
 PA (UYOR-) UNIT OREGON.
 PI Cuditt AB, Heim R, Otmo MF, Remington JS, Tsien RY;
 XX
 XX WPI; 1998-159454/14.
 DR
 XX
 XX
 PT Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS Claim 5; Page -; 120pp; English.
 XX
 CC This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX Sequence 238 AA:
 SO
 Query Match 99.8%; Score 1267; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4.1e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVNPILVELDGVNGHKRSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKRSVSGEGSDVYTKLTKICTTGKLPVMPPTL 60
 QY 61 VTFPSYGVQCFSRYPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVFEEDTLV 120
 DB 61 VTFPSYGVQCFSRYPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKTRAEVFEEDTLV 120
 QY 121 NRLEKIGIDKEGDNILGHKLEYNYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRLEKIGIDKEGDNILGHKLEYNYNSHNYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
 QY 181 YVOQNPILDGPVLLPDPNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YVOQNPILDGPVLLPDPNHYLSTQSALSKDPNEKRDMVLEFVTAAGITHGMDLYK 238
 RESULT 3
 AAM52330
 ID AAM52330 standard; Protein; 238 AA.
 XX
 AC AAM52330;
 XX
 DT 10-JUL-1998 (first entry)
 XX
 XX Engineered green fluorescent protein S65A.
 DE
 XX
 KM Green fluorescent protein; engineered fluorescent protein; mutain;

KM reporter molecule; immunological assay; protein-protein interaction;
 KM fluorescence resonance energy transfer system; FRET system.
 XX Synthetic.
 OS Aequorea victoria.
 XX Aequorea victoria.
 FH Key Location/Qualifiers
 FT Misc-difference 65 /Label= S65A
 FT /note= "Ser to Ala mutation"
 XX
 XX
 XX WO9806737-A1.
 XX
 XX 19-FEB-1998.
 XX
 XX 15-AUG-1997; 97WO-US14593.
 XX
 XX 30-AUG-1996; 96US-0706408.
 XX 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 XX (RECC) UNIV CALIFORNIA.
 XX (UYOR-) UNIV OREGON.
 XX
 XX Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 XX
 XX WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 XX
 XX Claim 5; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 XX
 XX Sequence 238 AA:
 SQ
 Query Match 99.8%; Score 1267; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 4, 1e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELTAVVPIIVELDGDVNGHKFSVSGEGEDVYTGKLTLCFTCTGKLPVWPPTL 60
 DB 1 MSKGEELTAVVPIIVELDGDVNGHKFSVSGEGEDVYTGKLTLCFTCTGKLPVWPPTL 60
 QY 61 VTTFEYGVQCSFRYRDHKKRHHDFKSSAMPEGVQOORTTFKRDGNYKTPAEKFGEDTLY 120
 DB 61 VTTFEYGVQCSFRYRDHKKRHHDFKSSAMPEGVQOORTTFKRDGNYKTPAEKFGEDTLY 120
 QY 121 NRLEKGIIDFKEDGNILGKLEYNYNSHNVYIMADKOKNGIKVNFKIRINIDGQVQLAD 180
 DB 121 NRLEKGIIDFKEDGNILGKLEYNYNSHNVYIMADKOKNGIKVNFKIRINIDGQVQLAD 180
 QY 181 YYQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITGHMDLYK 238
 DB 181 YYQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITGHMDLYK 238

DB 181 YYQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHMLLEFVTAAGITGHMDLYK 238
 RESULT 4
 ID AAM52333 standard; Protein: 238 AA.
 AC AAM52333;
 DT 10-JUL-1998 (first entry)
 XX
 XX Engineered green fluorescent protein Y66F.
 XX
 XX Green fluorescent protein; engineered fluorescent protein; mutetin;
 KM reporter molecule; immunological assay; protein-protein interaction;
 KM fluorescence resonance energy transfer system; FRET system.
 XX Synthetic.
 OS Aequorea victoria.
 XX Aequorea victoria.
 FH Key Location/Qualifiers
 FT Misc-difference 66 /Label= Y66F
 FT /note= "Tyr to Phe mutation"
 XX
 XX
 XX WO9806737-A1.
 XX
 XX 19-FEB-1998.
 XX
 XX 15-AUG-1997; 97WO-US14593.
 XX
 XX 30-AUG-1996; 96US-0706408.
 XX 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 XX (RECC) UNIV CALIFORNIA.
 XX (UYOR-) UNIV OREGON.
 XX
 XX Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 XX
 XX WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 XX
 XX Claim 5; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 XX
 XX Sequence 238 AA:
 SQ
 Query Match 99.7%; Score 1266; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 3, 3e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MSKGEELTAVVPIIVELDGVNGHKFSVSGEGEDVYTGKLTIKTCTGKLPVPMPTL 60
DB 1 MSKGEELTAVVPIIVELDGVNGHKFSVSGEGEDVYTGKLTIKTCTGKLPVPMPTL 60
QY 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKFECDTLV 120
DB 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKFECDTLV 120
QY 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHVLLEFYTAGITGMDLYK 238
DB 181 YVQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHVLLEFYTAGITGMDLYK 238

RESULT 5
AAW52315
ID AAW52315 standard; Protein: 238 AA.
XX
AC AAW52315;
XX
DE 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein.
XX
KW Green fluorescent protein; engineered fluorescent protein; muten;
KW reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Synthetic.
OS Aequorea victoria.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 65
FT /label= Gly, Thr, Ala, Leu, Cys, Val, Ile
XX
XX W09806737-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14593.
XX
XX 30-AUG-1996; 96US-0706408.
XX
XX 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AURORA BIOSCIENCES.
XX (REGC ) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON.
XX
XX Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX WPI: 1998-159454/14.
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
XX longer wavelength emission - used as markers for probes and as
XX components of fluorescent resonant energy transfer systems, also
XX related vectors and transformants
XX
XX Claim 2; Page -: 120pp; English.
XX
XX This sequence is an engineered Aequorea victoria green fluorescent
XX protein (GFP) of the invention, created from the sequence shown in
XX AAW52313. The GFP DNA was mutated to produce the nucleic acids (I) of
XX the invention, which encode functional, engineered fluorescent proteins
XX (II) having largely the same sequence as Aequorea green fluorescent
XX protein (GFP) but differing by at least the substitution T203X (X = H,
XX Y, W or F) and having fluorescent properties different from GFP. (II)
XX are useful as reporter molecules in immunological or hybridisation
XX assays, for monitoring proteins in cells and detecting induction of
XX transcription. They are also useful as components of FRET (fluorescence
XX resonance energy transfer) systems, e.g. for detecting protein-protein

```

```

CC interactions, cleavage of substrates and changes in potential across a
CC membrane, and further for making fluorescent substrates for protein
CC kinase. (II) are easily distinguished from known green and blue
CC fluorescent proteins, so allow simultaneous measurements of at least 2
CC proteins in a cell. They emit, and are excited, at relatively long
CC wavelengths where phototoxicity and auto-fluorescence are reduced.
XX
SQ Sequence 238 AA;
Query Match 99.7%; Score 1266; DB 19; Length 238;
Best Local Similarity 99.6%; Pred. No. 5.3e-123;
Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MSKGEELTAVVPIIVELDGVNGHKFSVSGEGEDVYTGKLTIKTCTGKLPVPMPTL 60
DB 1 MSKGEELTAVVPIIVELDGVNGHKFSVSGEGEDVYTGKLTIKTCTGKLPVPMPTL 60
QY 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKFECDTLV 120
DB 61 VTTFSYGVOCFSRYPDHKKRHDFFKSAMPEGYVOQRTIFFKDDGNYKTRAEVKFECDTLV 120
QY 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
DB 121 NRIELKGIDFEKEDGNILGHKLEYNNSHNHYIMADKQNGIKVNFKIRHNIEDGSVOLAD 180
QY 181 YVQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHVLLEFYTAGITGMDLYK 238
DB 181 YVQONTPIIDGPVLLPDNHYLSTQSALSKDPNEKRDHVLLEFYTAGITGMDLYK 238

RESULT 6
AAW52351
ID AAW52351 standard; Protein: 238 AA.
XX
AC AAW52351;
XX
DE 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein.
XX
KW Green fluorescent protein; engineered fluorescent protein; muten;
KW reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
XX Synthetic.
XX Aequorea victoria.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 224
FT /label= His, Asn, Gln, Thr, Phe, Trp, Tyr
XX
XX W09806737-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14593.
XX
XX 30-AUG-1996; 96US-0706408.
XX
XX 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AURORA BIOSCIENCES.
XX (REGC ) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON.
XX
XX Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX WPI: 1998-159454/14.
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
XX longer wavelength emission - used as markers for probes and as
XX components of fluorescent resonant energy transfer systems, also
XX related vectors and transformants

```


PS Claim 47; Page -: 120pp; English.

CC This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

SO Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITKFTTGTGKLPVWPTL 60
 DB 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITKFTTGTGKLPVWPTL 60

OY 61 VTFSYGYQCFSRYPDHMKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDLY 120
 DB 61 VTFSYGYQCFSRYPDHMKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDLY 120

OY 121 NRIELKGIDFEKEDGNILGHKLELYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFEKEDGNILGHKLELYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

OY 181 YVOONTPIIDGPVLLPDPNHYLSTQSALSKDPNEKRDNHVLLEFYTAAGITGMDELYK 238
 DB 181 YVOONTPIIDGPVLLPDPNHYLSTQSALSKDPNEKRDNHVLLEFYTAAGITGMDELYK 238

RESULT 7
 AAM52335
 ID AAM52335 standard; Protein: 238 AA.

AC AAM52335;
 DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

XX Green fluorescent protein; engineered fluorescent protein; mutin;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 OS Aequorea victoria.

XX Key Location/Qualifiers
 FT Misc-difference 42
 FT /Label- Cys, Phe, His, Trp, Tyr

XX WO9806737-A1.
 XX 19-FEB-1998.
 XX 15-AUG-1997; 97WO-US14593.
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX

PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC.) UNIV CALIFORNIA.
 PA (UOR-) UNIV OREGON.

PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI, 1998-159454/14.

XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 47; Page -: 120pp; English.

CC This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

SO Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITKFTTGTGKLPVWPTL 60
 DB 1 MSKGEELFTAVVPLIVELDGVNGHKFSVSGEGSDVYGGKLTITKFTTGTGKLPVWPTL 60

OY 61 VTFSYGYQCFSRYPDHMKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDLY 120
 DB 61 VTFSYGYQCFSRYPDHMKRHDFFKSAPEGYVOORTIFFDDGNYKTRAVEKFGDLY 120

OY 121 NRIELKGIDFEKEDGNILGHKLELYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKGIDFEKEDGNILGHKLELYNNSHNVIYIMADKONGIKVNFKIRHNIEDGSVOLAD 180

OY 181 YVOONTPIIDGPVLLPDPNHYLSTQSALSKDPNEKRDNHVLLEFYTAAGITGMDELYK 238
 DB 181 YVOONTPIIDGPVLLPDPNHYLSTQSALSKDPNEKRDNHVLLEFYTAAGITGMDELYK 238

RESULT 8
 AAM52336
 ID AAM52336 standard; Protein: 238 AA.

AC AAM52336;
 DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

XX Green fluorescent protein; engineered fluorescent protein; mutin;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 OS Aequorea victoria.

FH Key Location/Qualifiers
 FT Misc-difference 61 /Label= Phe, Tyr, His, Cys
 XX
 XX
 PN W09806737-A1.
 XX
 PD 19-FEB-1998.
 XX
 PF 15-AUG-1997; 97WO-US14593.
 XX
 PR 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI; 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 PS
 PS Claim 47; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA;
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVVPLIVELDDGVNGHKFSVSGEGEDVTYKLTFLKCTTGKLPVMPPTL 60
 Db 1 MSKGEELFTAVVPLIVELDDGVNGHKFSVSGEGEDVTYKLTFLKCTTGKLPVMPPTL 60
 QY 61 VTFPSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGNGYTRAEVFEEDTLV 120
 Db 61 VTFPSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGNGYTRAEVFEEDTLV 120
 QY 121 NRTELKGIIDFKEDGNILGHLEKLYNNSHNHYIMADKKNKGIKVFRIHNEEDSSVOLAD 180
 Db 121 NRTELKGIIDFKEDGNILGHLEKLYNNSHNHYIMADKKNKGIKVFRIHNEEDSSVOLAD 180
 QY 181 YYOQNTPIIDGPIVLLPDMHILSTQSALSKDPNEKRDMVLLFEPTAAGITGHMDELK 238
 Db 181 YYOQNTPIIDGPIVLLPDMHILSTQSALSKDPNEKRDMVLLFEPTAAGITGHMDELK 238
 RESULT 9
 AAM52337
 ID AAM52337 standard; Protein: 238 AA.
 XX

AC AAM52337;
 XX
 XX 10-JUL-1998 (first entry)
 DT
 XX
 XX Engineered green fluorescent protein.
 DE
 XX
 XX Green fluorescent protein; engineered fluorescent protein; mulein;
 KW reporter molecule; immunological assay; protein-protein interaction;
 KW fluorescence resonance energy transfer system; FRET system.
 XX
 XX Synthetic.
 OS Aequorea victoria.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 62 /Label= Ala, Val, Phe, Ser, Asp, Asn, Gln, Tyr, His, Cys
 FT
 FT
 XX W09806737-A1.
 XX
 XX 19-FEB-1998.
 XX
 XX 15-AUG-1997; 97WO-US14593.
 XX
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 XX (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 XX Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 PI
 DR WPI; 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 PS
 PS Claim 47; Page -: 120pp; English.
 XX
 XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA;
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSKGEELFTAVVPLIVELDDGVNGHKFSVSGEGEDVTYKLTFLKCTTGKLPVMPPTL 60
 Db 1 MSKGEELFTAVVPLIVELDDGVNGHKFSVSGEGEDVTYKLTFLKCTTGKLPVMPPTL 60
 QY 61 VTFPSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGNGYTRAEVFEEDTLV 120
 Db 61 VTFPSYGVQCFSRYPDMKRRHDFKFSAMPBGYVOQRTIFFKDGNGYTRAEVFEEDTLV 120

QY 121 NRIELKGIIDFKEDNGILGHKLEIYNNSHNYIMADKOKNGIKVFKIRHNIEDSGVOLAD 180
 Db 121 NRIELKGIIDFKEDNGILGHKLEIYNNSHNYIMADKOKNGIKVFKIRHNIEDSGVOLAD 180
 QY 181 YYQONTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238
 Db 181 YYQONTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238

RESULT 10

AAW52338
 ID AAW52338 standard; Protein: 238 AA.

AC AAW52338;
 XX 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; mutetin;
 KM Reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 OS Aequorea victoria.

FT Key Location/Qualifiers
 FT Misc-difference 68
 FT /Label- Phe, Tyr, His

PN WO9806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

XX (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;

XX WPI: 1998-159454/14.

XX Nucleic acid encoding mutant green fluorescent proteins having
 FT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAW52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit and are excited at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA;
 S0

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6,7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVVPIILVELDGVNGHKFSVSGEGDVTYGLTLKFTCTGKLPVMPPTL 60
 Db 1 MSKGEELFTAVVPIILVELDGVNGHKFSVSGEGDVTYGLTLKFTCTGKLPVMPPTL 60
 QY 61 VTTFSYGVOCSRPRPDHKKRDFPKSAMPEGYVOQRITFEFDGDNKYTRAVKREGDTLV 120
 Db 61 VTTFSYGVOCSRPRPDHKKRDFPKSAMPEGYVOQRITFEFDGDNKYTRAVKREGDTLV 120
 QY 121 NRIELKGIIDFKEDNGILGHKLEIYNNSHNYIMADKOKNGIKVFKIRHNIEDSGVOLAD 180
 Db 121 NRIELKGIIDFKEDNGILGHKLEIYNNSHNYIMADKOKNGIKVFKIRHNIEDSGVOLAD 180
 QY 181 YYQONTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238
 Db 181 YYQONTPIIDGPIVLLPDNHYLSTQSALSCKDNEKRDMHVLLEFYTAAGITGMDLYK 238

RESULT 11

AAW52344
 ID AAW52344 standard; Protein: 238 AA.

AC AAW52344;
 XX 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

KW Green fluorescent protein; engineered fluorescent protein; mutetin;
 KM Reporter molecule; immunological assay; protein-protein interaction;
 XX fluorescence resonance energy transfer system; FRET system.

OS Synthetic.
 OS Aequorea victoria.

FT Key Location/Qualifiers
 FT Misc-difference 150
 FT /Label- Phe, Tyr, His

PN WO9806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

XX (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;

XX WPI: 1998-159454/14.

XX Nucleic acid encoding mutant green fluorescent proteins having
 FT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAW52333. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,

Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6,7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGDVTYKLTIKFTCTGKLPVWPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGDVTYKLTIKFTCTGKLPVWPTL 60
 QY 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKYTRAVEREGDTLV 120
 DB 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKYTRAVEREGDTLV 120
 QY 121 NRIELKIDPKEDGNILGHKLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKIDPKEDGNILGHKLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 QY 181 YYOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YYOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 12

AAW52346
 ID AAW52346 standard; Protein: 238 AA.

AC AAW52346;

DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

XX Green fluorescent protein; engineered fluorescent protein; mutin;

KW reporter molecule; immunological assay; protein-protein interaction;

KM fluorescence resonance energy transfer system; FRET system.

XX Synthetic.

OS Aequorea victoria.

XX Key Location/Qualifiers

FT Misc-difference 167

FT /label= Phe, Tyr, His

PN WO9806737-A1.

PD 19-FEB-1998.

PE 15-AUG-1997; 97WO-US14593.

PR 30-AUG-1996; 96US-0706408.

PR 16-AUG-1996; 96US-0024050.

XX (AURO-) AURORA BIOSCIENCES.

PA (REGC) UNIV CALIFORNIA.

PA (UYOR-) UNIV OREGON.

XX Cubitt AB, Heim R, Ormo MF, Remington JS, Tsien RY;

XX WPI; 1998-159454/14.

PT Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants

PS Claim 47; Page -: 120pp; English.

XX This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAW52346. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.

XX Sequence 238 AA:

Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6,7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGDVTYKLTIKFTCTGKLPVWPTL 60
 DB 1 MSKGEELFTAVNPILVELDGVNGHKFSVSGEGDVTYKLTIKFTCTGKLPVWPTL 60
 QY 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKYTRAVEREGDTLV 120
 DB 61 VTFPSYGVQCFSRPDMKRRHDFPKSAMPEGYVOQRTIFPKDGNKYTRAVEREGDTLV 120
 QY 121 NRIELKIDPKEDGNILGHKLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 DB 121 NRIELKIDPKEDGNILGHKLEYNNSHNYIMADKKNKIKVFKIRHNIEDGSVOLAD 180
 QY 181 YYOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238
 DB 181 YYOQNPILDGPVLLPNNHYLSTOSALSCKPNEKRDMVLEFVTAAGITHGMDLYK 238

RESULT 13

AAW52349
 ID AAW52349 standard; Protein: 238 AA.

AC AAW52349;

DT 10-JUL-1998 (first entry)

DE Engineered green fluorescent protein.

XX Green fluorescent protein; engineered fluorescent protein; mutin;

KW reporter molecule; immunological assay; protein-protein interaction;

KM fluorescence resonance energy transfer system; FRET system.

XX Synthetic.

OS Aequorea victoria.

XX Key Location/Qualifiers

FT Misc-difference 220

FT /label= His, Asn, Gln, Thr

PN WO9806737-A1.

PD 19-FEB-1998.

PF 15-AUG-1997; 97WO-US14593.
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 XX Claim 47; Page -: 120pp; English.
 XX
 CC This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA:
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MSKGEELTFAVVPILVELDGVNGHKFVSGBEGDVTYTKLTKFICTTGKLPWPPTL 60
 DB 1 MSKGEELTFAVVPILVELDGVNGHKFVSGBEGDVTYTKLTKFICTTGKLPWPPTL 60
 QY 61 VTTFSYGVQCFSRYPDHNKRHDFFKSAMPEGVVQORTIFFKDDGNYKTRAEVKFGDPLY 120
 DB 61 VTTFSYGVQCFSRYPDHNKRHDFFKSAMPEGVVQORTIFFKDDGNYKTRAEVKFGDPLY 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKONKGIKVRNIRNIEDGSVQIAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKONKGIKVRNIRNIEDGSVQIAD 180
 QY 181 YVQONTPIIDGVPVLLPDNHYISTQSALSKDPNEKRDHMYLFEVTAAGITTHGMDLYK 238
 DB 181 YVQONTPIIDGVPVLLPDNHYISTQSALSKDPNEKRDHMYLFEVTAAGITTHGMDLYK 238
 RESULT 14
 AAM52324 ID AAM52324 standard; Protein: 238 AA.
 XX AAM52324;
 AC
 XX 10-JUL-1998 (first entry)
 XX
 DE Engineered green fluorescent protein Y66H.
 XX
 XX Green fluorescent protein; engineered fluorescent protein; mutain;
 KM reporter molecule; immunological assay; protein-protein interaction;

KM fluorescence resonance energy transfer system; FRET system.
 XX Synthetic.
 OS Aequorea victoria.
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 66
 FT /label= Y66H
 FT /note= "Tyr to His mutation"
 XX
 XX W09806737-A1.
 XX
 XX 19-FEB-1998.
 XX
 XX 15-AUG-1997; 97WO-US14593.
 XX
 XX 30-AUG-1996; 96US-0706408.
 PR 16-AUG-1996; 96US-0024050.
 XX
 PA (AURO-) AURORA BIOSCIENCES.
 PA (REGC) UNIV CALIFORNIA.
 PA (UYOR-) UNIV OREGON.
 XX
 PI Cubitt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
 DR WPI: 1998-159454/14.
 XX
 XX Nucleic acid encoding mutant green fluorescent proteins having
 PT longer wavelength emission - used as markers for probes and as
 PT components of fluorescent resonant energy transfer systems, also
 PT related vectors and transformants
 PS
 XX Claim 5; Page -: 120pp; English.
 XX
 CC This sequence is an engineered Aequorea victoria green fluorescent
 CC protein (GFP) of the invention, created from the sequence shown in
 CC AAM52313. The GFP DNA was mutated to produce the nucleic acids (I) of
 CC the invention, which encode functional, engineered fluorescent proteins
 CC (II) having largely the same sequence as Aequorea green fluorescent
 CC protein (GFP) but differing by at least the substitution T203X (X = H,
 CC Y, W or F) and having fluorescent properties different from GFP. (II)
 CC are useful as reporter molecules in immunological or hybridisation
 CC assays, for monitoring proteins in cells and detecting induction of
 CC transcription. They are also useful as components of FRET (fluorescence
 CC resonance energy transfer) systems, e.g. for detecting protein-protein
 CC interactions, cleavage of substrates and changes in potential across a
 CC membrane, and further for making fluorescent substrates for protein
 CC kinase. (II) are easily distinguished from known green and blue
 CC fluorescent proteins, so allow simultaneous measurements of at least 2
 CC proteins in a cell. They emit, and are excited, at relatively long
 CC wavelengths where phototoxicity and auto-fluorescence are reduced.
 CC
 XX
 SQ Sequence 238 AA:
 Query Match 99.6%; Score 1265; DB 19; Length 238;
 Best Local Similarity 99.6%; Pred. No. 6.7e-123;
 Matches 237; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSKGEELTFAVVPILVELDGVNGHKFVSGBEGDVTYTKLTKFICTTGKLPWPPTL 60
 DB 1 MSKGEELTFAVVPILVELDGVNGHKFVSGBEGDVTYTKLTKFICTTGKLPWPPTL 60
 QY 61 VTTFSYGVQCFSRYPDHNKRHDFFKSAMPEGVVQORTIFFKDDGNYKTRAEVKFGDPLY 120
 DB 61 VTTFSYGVQCFSRYPDHNKRHDFFKSAMPEGVVQORTIFFKDDGNYKTRAEVKFGDPLY 120
 QY 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKONKGIKVRNIRNIEDGSVQIAD 180
 DB 121 NRIELKGIDFKEDGNILGHKLEYNNSHVYIMADKONKGIKVRNIRNIEDGSVQIAD 180
 QY 181 YVQONTPIIDGVPVLLPDNHYISTQSALSKDPNEKRDHMYLFEVTAAGITTHGMDLYK 238
 DB 181 YVQONTPIIDGVPVLLPDNHYISTQSALSKDPNEKRDHMYLFEVTAAGITTHGMDLYK 238

RESULT 15
AAW52331
ID AAW52331 standard; Protein: 238 AA.
XX
AC AAW52331;
XX
DT 10-JUL-1998 (first entry)
XX
DE Engineered green fluorescent protein S65C.
XX
KW Green fluorescent protein; engineered fluorescent protein; mutain;
KW reporter molecule; immunological assay; protein-protein interaction;
KW fluorescence resonance energy transfer system; FRET system.
XX
OS Synthetic.
OS Aequorea victoria.
XX
FH Key Location/Qualifiers
FT Misc-difference 65 /label= S65C
FT /note= "Ser to Cys mutation"
XX
XX
PN M09806737-A1.
XX
XX 19-FEB-1998.
XX
XX 15-AUG-1997; 97WO-US14593.
XX
XX 30-AUG-1996; 96US-0706408.
XX 16-AUG-1996; 96US-0024050.
XX
XX (AURO-) AURORA BIOSCIENCES.
XX (REGC) UNIV CALIFORNIA.
XX (UYOR-) UNIV OREGON.
XX
XX Cuditt AB, Helm R, Ormo MF, Remington JS, Tsien RY;
XX
XX WPI: 1998-159454/14.
XX
XX
XX Nucleic acid encoding mutant green fluorescent proteins having
XX longer wavelength emission - used as markers for probes and as
XX components of fluorescent resonant energy transfer systems, also
XX related vectors and transforants
XX
XX
XX PS Claim 5; Page -: 120p; English.
XX
XX
XX This sequence is an engineered Aequorea victoria green fluorescent
XX protein (GFP) of the invention, created from the sequence shown in
XX AAW52313. The GFP DNA was mutated to produce the nucleic acids (I) of
XX the invention, which encode functional, engineered fluorescent proteins
XX (II) having largely the same sequence as Aequorea green fluorescent
XX protein (GFP) but differing by at least the substitution T203X (X = H,
XX Y, W or F) and having fluorescent properties different from GFP. (II)
XX are useful as reporter molecules in immunological or hybridisation
XX assays, for monitoring proteins in cells and detecting induction of
XX transcription. They are also useful as components of FRET (fluorescence
XX resonance energy transfer) systems, e.g. for detecting protein-protein
XX interactions, cleavage of substrates and changes in potential across a
XX membrane, and further for making fluorescent substrates for protein
XX kinase. (II) are easily distinguished from known green and blue
XX fluorescent proteins, so allow simultaneous measurements of at least 2
XX proteins in a cell. They emit, and are excited, at relatively long
XX wavelengths where phototoxicity and auto-fluorescence are reduced.
XX
XX
XX Sequence 238 AA;
XX
XX
XX Query Match 99.6%; Score 1265; DB 19; Length 238;
XX Best Local Similarity 99.6%; Pred. No. 6,7e-123;
XX Matches 237; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1 MSKGEEFTAVVPILVELDGDVNGHKFSVSGEGEDVTYGGKLTLEKICTTGKLPVMPPTL 60.

Db 1 MSKGEEFTAVVPILVELDGDVNGHKFSVSGEGEDVTYGGKLTLEKICTTGKLPVMPPTL 60
QY 61 VTTFSGYVOCPSRYPDHMKRRHDFKSAPEGYVOORTLEFDDGNYKTRAVKREGDPTV 120
Db 61 VTTFSGYVOCPSRYPDHMKRRHDFKSAPEGYVOORTLEFDDGNYKTRAVKREGDPTV 120
QY 121 NRLEKGIIDFKEDGNILGKLEYNYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
Db 121 NRLEKGIIDFKEDGNILGKLEYNYNNSHNYYIMADKOKNGIKVFKIRHNIEDGSVOLAD 180
QY 181 YVOQNTPIIDGPVLLPDNHYLSTOSALSADPNKRDHMYLLEFYTAAGITRGMDELTK 238
Db 181 YVOQNTPIIDGPVLLPDNHYLSTOSALSADPNKRDHMYLLEFYTAAGITRGMDELTK 238

Search completed: July 11, 2003, 17:56:46
Job time : 74 secs